

markovMSM: An R Package for Checking the Markov Condition in Multi-State Survival Data

by Gustavo Soutinho and Luís Meira-Machado

Abstract Multi-state models can be successfully used to describe processes in which an individual moves through a finite number of states in continuous time. These models allow a detailed view of the evolution or recovery of the process and can be used to study the effect of a vector of explanatory variables on the transition intensities or to obtain prediction probabilities of future events after a given event history. In both cases, before using these models, we have to evaluate whether the Markov assumption is tenable. This paper introduces the **markovMSM** package, a software application for R, which considers tests of the Markov assumption that are applicable to general multi-state models. Three approaches using existing methodology are considered: a simple method based on including covariates depending on the history; methods based on measuring the discrepancy of the non-Markov estimators of the transition probabilities to the Markovian Aalen-Johansen estimators; and, finally, methods that were developed by considering summaries from families of log-rank statistics where individuals are grouped by the state occupied by the process at a particular time point. The main functionalities of the **markovMSM** package are illustrated using real data examples.

1. Introduction

Multi-state models are commonly used to analyze complex longitudinal survival data involving multiple events of interest (Andersen et al., 1993; Hougaard, 2000; Putter et al., 2007; Meira-Machado et al., 2009; Meira-Machado and Sestelo, 2019). These models can be considered as a generalization of the survival process where survival is the ultimate outcome of interest but where information is available about intermediate events that individuals may experience during the study period. For instance, in some biomedical applications, besides the 'healthy' initial state and the absorbing 'dead' state, one may observe intermediate (transient) states based on health conditions (e.g., heart or lung disease), disease stages (e.g., stages of cancer, HIV infection or Alzheimer's), clinical symptoms (e.g., bleeding episodes), biological markers (e.g., CD4 T-lymphocyte cell counts; serum immunoglobulin levels) or they can represent a non-fatal complication in the course of the illness (e.g., cancer recurrence, transplantation, etc.). The number of states and transitions among them defines the complexity of the model, which is illustrated by a direct graph with rectangular boxes to represent possible states and with arrows between the states representing the allowed transitions. Among the multi-state models, the simplest is the mortality model for survival data (with only two states). The competing risks model (Andersen and Keiding, 2002; Putter et al., 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. The most common is the illness-death model, also known as the disability model, which comprises three states. In the irreversible version of this model (Figure 1), individuals start in the 'alive and disease-free' state and subsequently move either to the 'diseased' state or to the 'dead' state. Individuals in the 'diseased' state will eventually move to the 'dead' state without any possibility of recovery. Data that arises from the model depicted in Figure 1 has been termed as semi-competing risk data and can be seen as an emerging challenge for time-to-event data where two events of interest are studied: non-terminal (e.g., disease diagnosis) and terminal (e.g., death) events. The non-terminal event is observed only if it precedes the terminal event, which may occur before or after the non-terminal event. In this context, the terminal event is a competing risk for the intermediate state but not the contrary (Xu et al., 2010; Li and Peng, 2017; Nevo and Gorfine, 2020).

Multi-state models can be used to assess the effects of covariates on the transition intensities, but they can also be used to obtain conditional prediction probabilities of future events. From the transition intensities, it is possible to evaluate the probability of one event ahead, but the transition probabilities are the relevant predictive quantities since they consider several transitions ahead. The occupation probabilities, often used in practice, are particular examples of transition probabilities in which the time origin is taken as the current time. When the underlying stochastic process is Markov, an elegant theory links the transition intensities to the transition probabilities, leading to the Aalen-Johansen estimator (Aalen and Johansen, 1978). This assumption claims that given the present state, the future evolution of the process is independent of the states previously visited and the transition times among them; in other words, the history of the process is summarized by the current state occupied. The Aalen-Johansen estimator is adapted to censoring, and it benefits from the assumption

of Markovianity to get consistent estimates. When the multi-state model is non-Markov, this is no longer the case. In fact, there are many applications for which the Markov assumption may not be verified; for example, the arrival time at the current state of the process often influences the transition intensities, leading to non-Markov structures (Andersen et al., 2000, Andersen and Keiding, 2002). Although the Aalen-Johansen estimator may be used to consistently estimate occupation probabilities for non-Markov multi-state models (Datta and Satten, 2001), in general it provides biased estimators if the Markov assumption does not hold. To tackle this issue, Meira-Machado et al. (2006) introduced, for the first time, estimators that do not rely on the Markov assumption. These authors showed the practical superiority of their estimators relative to the Aalen-Johansen estimator in situations in which the Markov condition is strongly violated. However, this approach has the drawback of requiring that the support of the censoring distribution contain the support of the lifetime distribution, otherwise they only report valid estimators for truncated transition probabilities. Recently, alternative estimators that are consistent regardless of the Markov condition and the referred assumption on the censoring support were proposed by de Uña-Álvarez and Meira-Machado (2015). 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. The same idea of subsampling, combined with the Aalen-Johansen estimator, was later used by Putter and Spitoni (2018) to introduce new estimators which were termed landmark Aalen-Johansen.

In the multi-state setting, it is of practical interest to determine whether the Markov property holds within a particular data set. This information can be used to assess the validity of the transition probability estimates as well as the best option to perform inference for the transition intensities when one aims to relate the individual characteristics to the intensity rates through a multi-state regression model. This assumption is usually checked by including covariates depending on the history (Kay, 1986; Andersen et al., 2000; Andersen and Keiding, 2002). For the progressive illness-death model, for example, the Markov assumption is particularly relevant for modeling the death transition after disease and, consequently, assessing whether this transition rate is affected by the time in the previous state. Alternative methods, based on a local Kendall's tau measuring the future-past association along time, were proposed by Rodríguez-Girondo and de Uña-Álvarez (2012) and Rodríguez-Girondo and de Uña-Álvarez (2016). The method proposed by these authors can be used for three-state progressive and illness-death models, but the extension of this test to general multi-state models is not straightforward, and thus, flexible methods that may be used in general models are required. A very recent work by Titman and Putter (2020) considers new approaches to check this assumption. In one of these approaches, local and general tests have been developed by considering summaries from families of log-rank statistics where individuals are grouped by the state occupied at varying initial time s . Chiou et al. (2018) also considered an equivalent problem to testing Markovianity (in the progressive illness-model) but involving tests for dependent truncation. Making use of the landmark approach, Soutinho and Meira-Machado (2021) introduces new methods for checking the Markov assumption through a local and global test given by measuring the discrepancy between the Aalen-Johansen estimator and the landmark estimators (de Uña-Álvarez and Meira-Machado, 2015, Putter and Spitoni, 2018) that are free of the Markov condition.

There is software available for multi-state survival analysis. A comprehensive list of the available packages in the Comprehensive R Archive Network (CRAN) can be seen in the CRAN task view 'Survival Analysis' (Allignol and Latouche, 2022). In particular, in the context of multi-state models, **mstate** package provides functions for estimating hazards and probabilities, possibly depending on covariates. This package also permits checking the Markov assumption from families of log-rank statistics where individuals are grouped by the state occupied at different times (Titman and Putter, 2020). The **msm** package can be used to fit a Markov model with any number of states and any pattern of transitions to panel data. It includes several extensions, such as hidden Markov models and models whose transition intensities vary with individual-specific or time-varying covariates (Jackson, 2011). **survidm** package provides estimates of predictive probabilities, such as the transition probabilities, occupation probabilities, cumulative incidence functions, and waiting (sojourn) time distributions. It also allows one to perform multi-state regression for transition intensities as well as to implement global tests for the Markov assumption in illness-death models (Soutinho et al., 2021). Other examples are the **SemiMarkov** package that can be used to fit semi-Markov multistate models to longitudinal data proposed by Król and Saint-Pierre (2015) and the **frailtyEM** which contains functions for fitting shared frailty models with a semi-parametric baseline hazard with the Expectation-Maximization algorithm (Balan and Putter, 2019). Multi-state regression can be performed using several of the aforementioned packages by decoupling the whole process into various survival models, by fitting separate intensities to all permitted transitions using, for example, semi-parametric Cox proportional hazard regression models, while making appropriate adjustments to the risk set. To this end, the **JM** and **joiner** packages, proposed by Rizopoulos (2010) and Philipson et al. (2018), can be useful to estimate the parameters for the joint modeling of longitudinal and time-to-event data. In this context, we can also refer to the **joinerML** which extends the inference of the joint model to the case

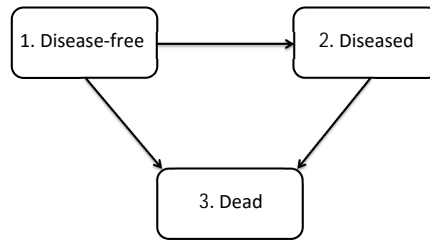


Figure 1: The progressive illness-death model with three states and three irreversible transitions. ‘Disease-free’ correspond to the initial state, ‘Diseased’ the intermediate state and the absorbing state ‘dead’.

of multiple continuous longitudinal measures (Hickey et al., 2018).

The organization of this paper is as follows: In Section 2, we present a brief introduction of the notation and mathematical background of existing methods for testing the Markov assumption. In Section 3, we introduce the `markovMSM` package, a software application for R that performs these local and global tests in the multi-state models described in Soutinho and Meira-Machado (2021) using three real data examples. Finally, the main conclusions are reported in Section 4.

2. Background and notation

This section contains a brief description of the mathematical background underlying the `markovMSM` package. Further details on the proposed methods can be found in Soutinho and Meira-Machado (2021).

Multi-state models and transition probabilities

A multi-state model $(X(t), t \in [0, \infty))$ is a model for a time continuous stochastic process. These models provide a relevant modeling framework to deal with complex longitudinal survival data since they allow the representation of the movement of the individuals among a set of finite states space $\mathcal{S} = 1, \dots, K$. They can be fully characterized through the transition probabilities, for two states h, j and two time points $s < t$, that are given by the conditional probability $p_{hj}(s, t) = P(X(t) = j | X(s) = h, \mathcal{H}_{s-})$, where \mathcal{H}_{s-} denotes the history of the process up to time s . This history can represent information about the process, such as the states previously visited, transition times, etc. Under the Markov assumption, $P(X(t) = j | X(s) = h, X(u) = y) = P(X(t) = j | X(s) = h)$, for any $0 \leq u < s$ and $y \in \mathcal{S}$, 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.

Without loss of generality and for the purpose of simplicity, let us consider the progressive illness-death model, with state space $\{1, 2, 3\}$, given by the initial State (e.g., ‘alive and disease-free’; State 1), the intermediate state (e.g., ‘diseased’; State 2), and the absorbing state (e.g., ‘dead’; State 3). For this type of model, there are five different transition probabilities to be considered that can be obtained as follows (Meira-Machado and Sestelo, 2019):

$$p_{11}(s, t) = \exp \left(- \int_s^t (\lambda_{12}(u) + \lambda_{13}(u)) du \right)$$

$$p_{22}(s, t | t_{12}) = \exp \left(- \int_s^t \lambda_{23}(u, t_{12}) du \right)$$

$$p_{12}(s, t) = \int_s^t p_{11}(s, u-) \lambda_{12}(u) p_{22}(u, t | u) du.$$

Here, $p_{22}(s, t | t_{12})$ denotes the transition probability p_{22} conditionally on a particular entry time t_{12} and $\lambda_{hj}(\cdot)$ correspond to the transition intensities that are the instantaneous hazards for movement from one state (h) to another (j). If the process is Markov, $h_{23}(t, t_{12}) = h_{23}(t)$ and $p_{22}(s, t | t_{12}) = p_{22}(s, t)$. The two other transition probabilities, $p_{13}(s, t)$ and $p_{23}(s, t)$, can be estimated from the two obvious relations that exist in the progressive illness-death model: $p_{11}(s, t) + p_{12}(s, t) + p_{13}(s, t) = 1$ and $p_{22}(s, t) + p_{23}(s, t) = 1$. Expressions for general models are not possible. Since the prognosis for an individual in the intermediate state may be influenced by the subject’s specific arrival time, the illness-death model is not necessarily Markovian.

The estimation of the transition probabilities is a major goal in multi-state models since they allow for long-term predictions. A common nonparametric method to estimate these quantities is the Aalen-Johansen (AJ) estimator, (Aalen and Johansen, 1978) which can be obtained as the product-integral of the Nelson-Aalen estimators for the cumulative transition intensities (Andersen et al., 1993). Explicit formulae of the Aalen-Johansen estimator for the illness-death model can be found in Borgan (2005). When the multi-state process is Markovian, the AJ estimator provides consistent estimates of the transition probabilities, but this is no longer the case if the process is non-Markov. To avoid this problem, de Uña-Álvarez and Meira-Machado (2015) uses the idea of subsampling, also referred to as landmarking (Van Houwelingen, 2007), which is based on (differences between) Kaplan-Meier estimators derived from a subset of the data consisting of all subjects observed to be in the given state at the given time. For the specific case of the illness-death model, given the time point s , to estimate $p_{1j}(s, t)$ for $j = 1, 2, 3$ the landmark analysis is restricted to the individuals observed in State 1 at time s ; whereas, to estimate $p_{2j}(s, t)$, $j = 2, 3$, the landmark analysis proceeds from the sample restricted to the individuals observed in State 2 at time s . Further details on the formulae and proofs can be seen in de Uña-Álvarez and Meira-Machado (2015). The subsampling approach was later used by Putter and Spitoni (2018) to derive a landmark Aalen-Johansen estimator (LMAJ) 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) for which consistency has already been proved in multi-state models that are not necessarily Markov (Datta and Satten, 2001). It is worth mentioning that, in the illness-death model, the landmark Aalen-Johansen estimators (LMAJ) and the landmark estimators (LM) of the transition probabilities have shown similar performances. However, the former methods (LMAJ) can be used in general multi-state models, which can be considered an advantage.

From now on, we will use the abbreviation AJ for the Aalen-Johansen estimator, LM for the LandMark estimator proposed by de Uña-Álvarez and Meira-Machado (2015), and LMAJ for the LandMark Aalen-Johansen estimator (Putter and Spitoni, 2018).

Tests for the Markov assumption

Modeling particular transition intensities

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; Andersen et al., 2000; Andersen and Keiding, 2002). In the progressive illness-death model, the Markov condition is particularly relevant for the transition from the intermediate state ‘disease’ (State 2) to ‘death’ (State 3). Under this model, we can examine whether the time spent in the initial state (State 1) is important in the transition from the disease state to death or not. For doing that, let $\lambda_{23}(t)$ denote the hazard function of T for those individuals going from State 2 to State 3. Fitting a Cox model $\lambda_{23}(t | Z) = \lambda_{23,0}(t) \exp(\beta Z)$, where $\lambda_{23,0}$ is the baseline hazard, Z the ‘time spent in the healthy (initial) state’ and β a regression parameter, we now need to test the null hypothesis, if the transition rate from the disease state into death is unaffected by the time spent in the healthy state, $H_0 : \beta = 0$, against the general alternative, $H_1 : \beta \neq 0$. This ‘global’ test, based on the semiparametric Cox proportional hazards model, assumes proportional hazards and a linear effect on the hazard for the covariate. When either of these two assumptions is not fulfilled, the test may not be able to detect the lack of Markovianity.

Since the landmark methods (LM) for estimating the transition probabilities proposed by de Uña-Álvarez and Meira-Machado (2015), and (LMAJ) by Putter and Spitoni (2018) are free of the Markov assumption, they can be used as a basis for the construction of tests for Markovianity. These ideas were used by Soutinho and Meira-Machado (2021) to introduce local and global tests for Markovianity based on the discrepancy of the landmark estimators to Markovian Aalen-Johansen estimators (AJ). Subsampling, also known as landmarking, has been also used by Titman and Putter (2020) to introduce a general test based on summaries from families of log-rank statistics where individuals are grouped by the state they occupied at a given (landmark) time.

AUC Local Test

Soutinho and Meira-Machado (2021) introduce a local test based on the area under curves, AUC, given by the estimated transition probabilities, for a fixed time $s > 0$, that can be used for a general multi-state model for checking the Markovianity of the process. They propose the following test statistic that is based on the difference between the area under the estimated transition probability curve for the non-Markov LMAJ estimator and the AJ estimator, $U = \int_s^\tau (\hat{p}_{hj}^{\text{LMAJ}}(s, u) - \hat{p}_{hj}^{\text{AJ}}(s, u)) du$, where τ is the upper bound of the support of T . When the process is Markovian, it is expected that the test statistic will be close to zero. The Markov assumption becomes less likely as the test statistic

gets further away from zero in either direction. Because of censoring, both estimators (LMAJ and AJ) may reveal high variability in the right tail, which may inflate the test statistic. In addition to this issue, since landmarking is based on reduced data, the maximum point for which the LMAJ transition probability estimate is strictly defined may be lower than the maximum point for AJ. To address these issues, we propose that when computing U , the minimum between the upper bound for which LMAJ is defined and the 90th percentile of the total time for the upper limit in the integral that defines the test statistic be used.

In the progressive illness-death model, besides the transition probability $\hat{p}_{23}(s, t)$, also $\hat{p}_{12}(s, t)$ can be used to test the Markov assumption. For general multi-state models, one can use transitions depending on history. In fact, if the goal is to decide which estimator is the most appropriate to use to estimate a specific transition probability $p_{hij}(s, t)$, then the test statistic should be the one based on that same transition probability.

To approximate the distributions of the test statistic, bootstrap methods with a large number of resamples, M , are used. We generate M bootstrap samples, and for each sample, we calculate the test statistic U^* which corresponds to the difference between the areas of the AJ and the LMAJ estimators. Finally, the variance $\sigma_{(U^*)}^2$ is obtained from the M values of U^* . Then, according to large sample asymptotic distribution theory, when M , the number of replicates, goes to infinity, we have the following statistic distributed approximately as a standard normal distribution with a mean of 0 and variance of 1: $V = (U - 0) / \sigma_{(U^*)} \sim N(0, 1)$. The null hypothesis will be rejected if $V > v_{(1-\alpha/2)}$ or $V < v_{(\alpha/2)}$, where $v_{(\alpha/2)}$ and $v_{(1-\alpha/2)}$ denote the $\alpha/2$ and $1 - \alpha/2$ percentiles, respectively, of a normal distribution with a mean of 0 and variance of 1.

AUC Global Test

Soutinho and Meira-Machado (2021) also introduces a global test, which can be achieved by combining the results obtained from local tests over different times. The testing procedure used here involves the following steps:

1. Using the original sample of the illness-death model, obtain the percentiles 5, 10, 20, 30, and 40 of the times spent by all individuals in State 1, the so-called waiting (sojourn) time. For general multi-state models, we recommend the use of the same percentiles of the subject's specific arrival time at the corresponding state.
2. For each of the values s obtained in Step 1, obtain the probability values for the local method as explained before.
3. Obtain the mean of the probability values for each closest pair; i.e., the mean of the probability values of the following pairs of percentiles: (5, 10), (10, 20), (20, 30) and (30, 40).
4. Finally, the p -value for the global test is equal to the minimum between the four probability values obtained in Step 3. In the case of this probability value being below a particular significance value, for example 0.05, we reject the null hypothesis of the process being Markovian.

Step 1 considers a global test based on local tests computed at low percentiles of subject-specific arrival times at the corresponding state. This is based on our experience that the failure of Markovianity often occurs at small transition times. Besides the hypothesis tests proposed above, we also suggest graphical local tests that can be used to check the Markov assumption in the illness-death model as well as for more complex multi-state models, possibly with reversible transitions between states. These graphical tests can be used to validate the default values proposed in Step 1 or to propose alternative values for which a discrepancy between the two methods (LMAJ and AJ) is more evident.

The procedure described in Step 3 can be used to ensure that there is a discrepancy between the two estimated curves over a large range of time values.

Log-rank test

The construction of the landmark Aalen-Johansen estimators (LMAJ) also motivated the development of new tests of the Markov assumption (Titman and Putter, 2020). Assume initially that interest lies in assessing the validity of transition probability estimates from a specific start point time s and a given state h , i.e. $p_{hij}(s, t)$. Under the landmark approach, the estimators of the transition probabilities should be derived from the subset of the data consisting of all subjects observed to be in the state h at time s , $S = \{i : X_i(s) = j, Y_i(s) = 1\}$ where $Y_i(s)$ is the at risk indicator of the process (de Uña-Álvarez and Meira-Machado, 2015; Putter and Spitoni, 2018). On the other hand, the Aalen-Johansen estimator would use also the set of subjects in $S^c = \{i : X_i(s) \neq j, Y_i(s) = 1\}$. Besides, under a Markov process, the transition intensities of the process for $t > s$ will be the same in the two groups. This motivated Titman and Putter (2020) to propose a local test for the Markov assumption that lies in

| Function | Description |
|-------------------|---|
| PHM.test | Performs a global test for the Markov assumption based on the Cox PH model. |
| AUC.test | Performs global and local tests for the Markov assumption using the Area Under Curves (AUC) method. |
| LR.test | Log-rank based test for the validity of the Markov assumption. |
| plot.markovMSM | Plot for an object of class markovMSM. |
| eventsMSM | Counts the number of observed transitions in the multi-state model. |
| prepMSM | Prepares the data set for multi-state modeling in a long format from a data set in wide format. |
| transMatMSM | Define transition matrix for a multi-state model. |
| print.markovMSM | Print for an object of class markovMSM. |
| summary.markovMSM | Summary for an object of class markovMSM. |

Table 1: Summary of functions in the `markovMSM` package.

testing $H_0 : \lambda(t|X(s) = h) = \lambda(t|X(s) \neq h)$ for $t \geq s$ versus a general alternative. The proposed test is constructed using several log-rank statistics grouped by membership in S . [Titman and Putter \(2020\)](#) also propose a global test **through** a grid of times in which, for each s time, local tests are computed. In this paper, the grid of values is given by the percentile times.

The same authors pointed out that the use of the individual supremum or integrated statistics may be combined to provide an overall test of the dependence of the transition intensities on occupation in state j at previous times.

3. Application of the markovMSM

This section offers the guidelines for the use of the `markovMSM` package with the R statistical program ([R Core Team, 2021](#)). To this end, a description of the functions of the package is illustrated using three real data sets. The first one involves data from a clinical trial on colon cancer modeled using the progressive illness-death model ([Moertel et al., 1995](#)). Extensions to progressive processes beyond the three-state illness-death model are **discussed** using data from the European Group for Blood and Marrow Transplantation (EBMT) ([Putter et al., 2007](#)). Finally, we use data from a study with liver cirrhosis patients subjected to prednisone treatment ([Andersen et al., 1993](#)). The package comprises seven main functions, which are briefly summarized in Table 1.

Data manipulation

In this section, we reanalyze data from a large clinical trial on Duke's stage III patients affected by colon cancer who underwent curative surgery for colorectal cancer ([Moertel et al., 1990](#)). Of the 929 patients, 468 developed a recurrence, and among these, 414 died; 38 died without a recurrence. The remaining 423 patients were alive and disease-free up to the end of the follow-up. Besides the two event times (time to recurrence and time to death) and the corresponding indicator statuses, a vector of covariates including age, sex, number of lymph nodes, and extent of local spread is also available. Below is an excerpt of the data set with one row per individual. Individuals were chosen in order to represent all possible combinations of movements among the three states.

```
> library(markovMSM)
> data("colonMSM")
> db_wide <- colonMSM
> head(db_wide[c(1:2, 16, 21), 1:11])
```

| | time1 | event1 | Stime | event | rx | sex | age | obstruct | perfor | adhere | nodes |
|----|-------|--------|-------|-------|---------|-----|-----|----------|--------|--------|-------|
| 1 | 968 | 1 | 1521 | 1 | Lev+5FU | 1 | 43 | 0 | 0 | 0 | 5 |
| 2 | 3087 | 0 | 3087 | 0 | Lev+5FU | 1 | 63 | 0 | 0 | 0 | 1 |
| 16 | 1323 | 1 | 3214 | 0 | Obs | 1 | 68 | 0 | 0 | 0 | 1 |
| 21 | 2789 | 0 | 2789 | 1 | Obs | 1 | 64 | 1 | 0 | 0 | 1 |

The four initial variables describe the movement of the patients among the three states of the illness-death model: `time1` denote the time measured in days from surgery to recurrence, whereas `Stime` is the total time or the time to death or censoring; `event1` and `event` denote the corresponding status/censoring indicator (1 for an event and 0 for censoring). Patient 1 had a recurrence after 968 days (i.e., observed a transition from the initial state to the intermediate state) and then died after 1521

days in the study. Patient 2 remains alive and without recurrence at the end of follow-up (event1 = 0 and event = 0). The two event times are equal in these cases. The patient represented in the third line had a recurrence after 1323 days but remained alive at the end of the follow-up (i.e., in State 2). Finally, the patient represented in the last line died after 2789 days in the study without experiencing a recurrence.

As the original data set is in the wide format, the next step to implementing the proposed methods will be to convert the data into a long format, which is given by one line for each transition for which a subject is at risk. This can be done using functions `transMatMSM` and `prepMSM`. `transMatMSM` function defines the transition matrices, revealing which transitions are possible, whereas `prepMSM` provides a new dataset in a long format for which each row will correspond to a transition for which a patient is at risk. For the progressive illness-death model, these two functions are used as follows:

```
> positions <- list(c(2, 3), c(3), c())
> state.names <- c("Alive", "Rec", "Death")
> tmat <- transMatMSM(positions, state.names)
> tmat
      to
from  Alive Rec Death
  Alive    NA  1    2
   Rec    NA NA    3
  Death    NA NA    NA

> timesNames <- c(NA, "time1", "Stime")
> status <- c(NA, "event1", "event")
> trans <- tmat
> db_long <- prepMSM(data = db_wide, trans, timesNames, status)
> db_long[1:10,]
```

| | id | from | to | trans | Tstart | Tstop | time | status |
|----|----|------|----|-------|--------|-------|------|--------|
| 1 | 1 | 1 | 2 | 1 | 0 | 968 | 968 | 1 |
| 2 | 1 | 1 | 3 | 2 | 0 | 968 | 968 | 0 |
| 3 | 1 | 2 | 3 | 3 | 968 | 1521 | 553 | 1 |
| 4 | 2 | 1 | 2 | 1 | 0 | 3087 | 3087 | 0 |
| 5 | 2 | 1 | 3 | 2 | 0 | 3087 | 3087 | 0 |
| 6 | 3 | 1 | 2 | 1 | 0 | 542 | 542 | 1 |
| 7 | 3 | 1 | 3 | 2 | 0 | 542 | 542 | 0 |
| 8 | 3 | 2 | 3 | 3 | 542 | 963 | 421 | 1 |
| 9 | 4 | 1 | 2 | 1 | 0 | 245 | 245 | 1 |
| 10 | 4 | 1 | 3 | 2 | 0 | 245 | 245 | 0 |

Finally, in terms of manipulation of data, a useful function is `eventsMSM` since it allows one to summarise the number of transitions among states and their percentages:

```
> eventsMSM(db_long)

$Frequencies
      to
from  Alive Rec Death no event total entering
  Alive    0 468   38    423          929
   Rec    0  0  414    54          468
  Death    0  0    0   452          452

$Proportions
      to
from  Alive      Rec      Death no event
  Alive 0.0000000 0.5037675 0.0409042 0.4553283
   Rec  0.0000000 0.0000000 0.8846154 0.1153846
  Death 0.0000000 0.0000000 0.0000000 1.0000000
```

Methods for testing the Markov condition in the illness-death model

The illness-death model is a special multi-state model with several applications in the biomedical literature. Many time-to-event data sets with multiple end points can be reduced to this generic

structure. Unlike the survival or competing risk models, this model is not necessarily Markovian since the prognosis for an individual in the intermediate state may be influenced by the subject's specific arrival time. Traditionally, the Markov assumption is checked by including covariates depending on the history. In the particular case of the colon cancer data set, we are interested in assessing if the transition rate from the recurrence state into death is unaffected by the time spent in the previous state. This global test for the Markov assumption can be done using the function `PHM.test`. A brief description of the arguments of this function is shown in Table 2. Results for this global test of our data indicated that the effect of the time spent in State 1 is not significant (p -value of 0.154), revealing no evidence against the Markov model for the colon data. The corresponding input codes are the following:

```
> res <- PHM.test(data = db_long, from = 2, to = 3)
[1] 0.1543195
> res
$p.value
[1] 0.1543195
$from
[1] 2
$to
[1] 3
```

| Argument | Description |
|----------|--|
| data | A data frame in the long format containing the subject id; from corresponding to the starting state; the receiving state, to; the transition number, trans; the starting time of the transition given by Tstart; the stopping time of the transition, Tstop, and status (for the status variable, with 1 indicating an event (transition), 0 a censoring). |
| from | The starting state of the transition to check the Markov condition. |
| to | The last state of the considered transition to check the Markov condition. |

Table 2: Summary of the arguments of the function `PHM.test`.

In the `markovMSM` package the local test proposed in Section 2.2.2 is performed using function `AUC.test`, through argument `type = 'local'`. A summary of the arguments of this function is presented in Table 3.

The input commands to perform the AUC local test, for a fixed time $s = 180$ and transitions $1 \rightarrow 2$ and $1 \rightarrow 3$ are the following:

```
> set.seed(1234)
> res2 <- AUC.test(data = db_long, times = 180, from = 1, to = 3, type = 'local',
                  replicas = 100, tmat = tmat)
> res2$localTest
      s    1->1    1->2    1->3
1 180 0.2902191 0.002982042 0.002992007
```

As a result, the function `AUC.test` returns the probability values for all attainable transitions from the initial state. To obtain the same local test for transition $2 \rightarrow 3$, we only need to put 2 in the parameter `from` as follows:

```
> set.seed(1234)
> res3 <- AUC.test(data = db_long, times = 180, from = 2, to = 3, type = 'local',
                  replicas = 100, tmat = tmat)
> res3$localTest
      s    2->2    2->3
1 180 0.02547708 0.04816232
```

Results reveal a possible failure of the Markov assumption with low probability values for transitions $1 \rightarrow 2$ and $2 \rightarrow 3$ for $s = 180$ (less than 5%). These findings agree with the results depicted in Figure 2, which reports the estimated transition probabilities. In fact, we can observe departures between the two Markov-free estimators (LM and LMAJ) and the Aalen-Johansen estimator (AJ) revealing a possible failure of the Markov assumption. The input commands to obtain the plots shown in Figure 2 are the following:

```
> plot(res2, to = 2, axis.scale = c(0,0.25), difP = FALSE) #Figure 2 (Left column)
> plot(res3, to = 3, axis.scale = c(0,1), difP = FALSE) # Figure 2 (Right column)
```


| Argument | Description |
|-------------|---|
| data | A data frame in the long format containing the subject id; from corresponding to the starting state; the receiving state, to; the transition number, trans; the starting time of the transition given by Tstart; the stopping time of the transition, Tstop, and status for the status variable, with 1 indicating an event (transition), 0 a censoring. |
| from | The starting state of the transition probabilities. |
| to | The last receiving state considered for the estimation of the transition probabilities. All the probabilities among the first and the last states are also computed. |
| type | Type of test for checking the Markov condition: local or global. By default type='global'. |
| times | For the local test, times represents the starting times of the transition probabilities. In case of a global test, the argument is given by times between the minimum time and the third quartile times used in the formula of this test. Default to NULL. |
| quantiles | Quantiles used in the formula of the global test for the AUC methods. |
| tmat | The transition matrix for multi-state model. |
| replicas | Number of bootstrap samples to standardization of the T-statistic given by the difference of the areas of AJ and LMAJ transition probabilities estimates. |
| limit | Percentile of the event time used as the upper bound for the computation of the AUC-based test. |
| positions | List of possible transitions; x[[i]] consists of a vector of state numbers reachable from state i. |
| namesStates | A character vector containing the names of either the competing risks or the states in the multi-state model specified by the competing risks or illness-death model. names should have the same length as the list x (for transMat), or either K or K + 1 (for trans.comprisk), or 3 (for trans.illdeath). |
| timesNames | Either 1) a matrix or data frame of dimension $n \times S$ (n being the number of individuals and S the number of states in the multi-state model), containing the times at which the states are visited or last follow-up time, or 2) a character vector of length S containing the column names indicating these times. In the latter cases, some elements of time may be NA |
| status | Either 1) a matrix or data frame of dimension $n \times S$, containing, for each of the states, event indicators taking the value 1 if the state is visited or 0 if it is not (censored), or 2) a character vector of length S containing the column names indicating these status variables. In the latter cases, some elements of status may be NA |

Table 3: Summary of the arguments of function AUC.test.

Putting the parameter `difP=TRUE`, we can also obtain the discrepancy between the Aalen-Johansen estimator (Markovian) and the landmark non-Markovian estimator (LMAJ), for $p_{12}(s, t)$ and $p_{23}(s, t)$, for $s = 180$, measured through $D_{hj} = \hat{p}_{hj}^{AJ}(s, t) - \hat{p}_{hj}^{LMAJ}(s, t)$, $h = 1, 2$, $j = h + 1$. The 95% point-wise confidence limits were obtained using a simple bootstrap (Figure 3). The corresponding input commands in this case are the following:

```
> plot(res2, to = 2, axis.scale = c(-0.03, 0.03), difP = TRUE) #Figure 3 (Left column)
> plot(res3, to = 3, axis.scale = c(-0.30, 0.10), difP = TRUE) #Figure 3 (Right column)
```

The plots in Figure 3 can be thought of as graphical local tests for the Markov assumption. As expected, in both cases they reveal differences between the two methods for $s = 180$ since they show a clear deviation with respect to the horizontal line $y = 0$. From these, one gets some (graphical) evidence of the lack of Markovianity of the underlying process beyond half a year after surgery. Therefore, for this specific time, the application of the Aalen-Johansen method may not be recommended here, due to possible biases.

In Section 2, a global test was also introduced that combines the probability values of the local test over different times (given by the percentiles of the sojourn time in State 1). In the [markovMSM](#) package, this can be obtained using the function `AUC.test`. The arguments for this function are described in Table 3. Some examples of how to perform the proposed AUC global test are shown in the following input commands, in which we consider the default percentiles in the argument

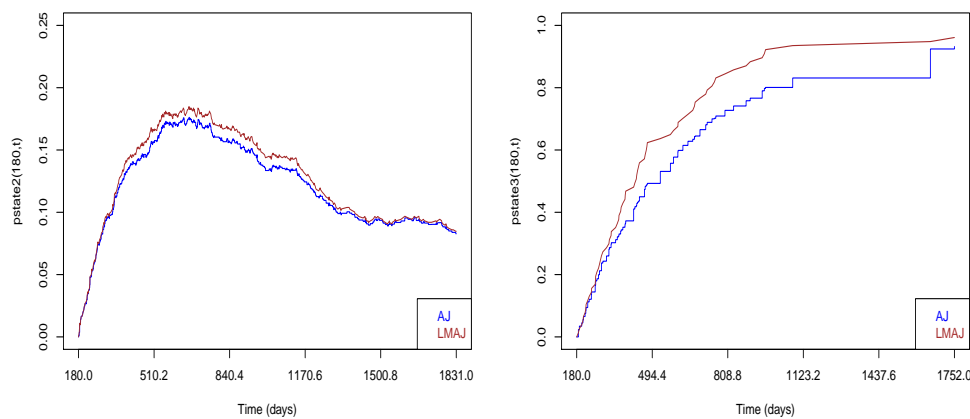


Figure 2: Estimated transition probabilities for the Aalen-Johansen (AJ) and Markov-free estimators (landmark and landmark Aalen-Johansen) using data from a colon cancer for $s = 180$. The displayed curves reveal a possible failure of the Markov assumption due to the departures between the LM and AJ estimates.

quantiles. Depending on the number of replicas, the implementation of the `AUC.test` function can be computationally time-consuming.

```
> set.seed(1234)
> res4 <- AUC.test(data = db_long, from = 1, to = 3, type = 'global', replicas = 100,
                   tmat = tmat)

> round(res4$globalTest,3)
  1->1  1->2  1->3
1 0.067 0.012 0.012

> set.seed(1234)
> res5 <- update(res4, from = 2)
> round(res5$globalTest,3)
  2->3
1 0.006
```

Results reported by the first command lines provide the probability values for the global test based on the AUC for the three transitions leaving State 1 (i.e., $1 \rightarrow 1$, $1 \rightarrow 2$ and $1 \rightarrow 3$). As expected, a higher probability value was obtained for transition $1 \rightarrow 1$ than the two remaining transitions, revealing evidence against the Markov condition. Results reported in the second set of input commands agree with previous findings, reporting a probability value of 0.006 for $2 \rightarrow 3$. Among the objects saved by this function, `AUC.test` displays the probability values for each percentile time (default to 5, 10, 20, 30, and 40) through the following codes:

```
> round(res4$localTest,3)
   s  1->1  1->2  1->3
1 102.4 0.978 0.081 0.081
2 173.0 0.118 0.025 0.025
3 290.6 0.015 0.000 0.000
4 469.2 0.679 0.056 0.056
5 726.8 0.635 0.176 0.176

> round(res5$localTest,3)
   s  2->2  2->3
1 102.4 0.015 0.015
2 173.0 0.011 0.011
3 290.6 0.000 0.000
4 469.2 0.050 0.050
5 726.8 0.156 0.156
```

These outputs show, for instance, that the probability value of the AUC local test reveal the fail of

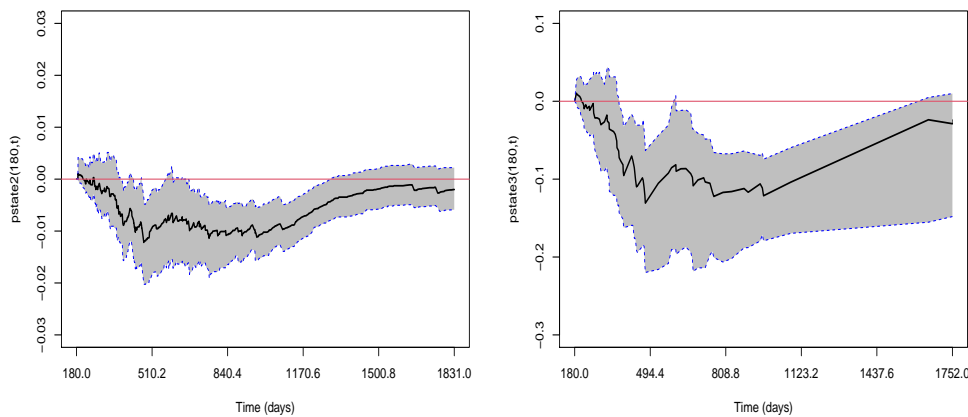


Figure 3: A local graphical test for the Markov condition for colon cancer data, for $s = 180$, based on the discrepancy between the Aalen-Johansen estimator (Markovian) and the Markov-free estimator (LMAJ) for each time from s . Bootstrap resamples were used to determine the 95% confidence interval.

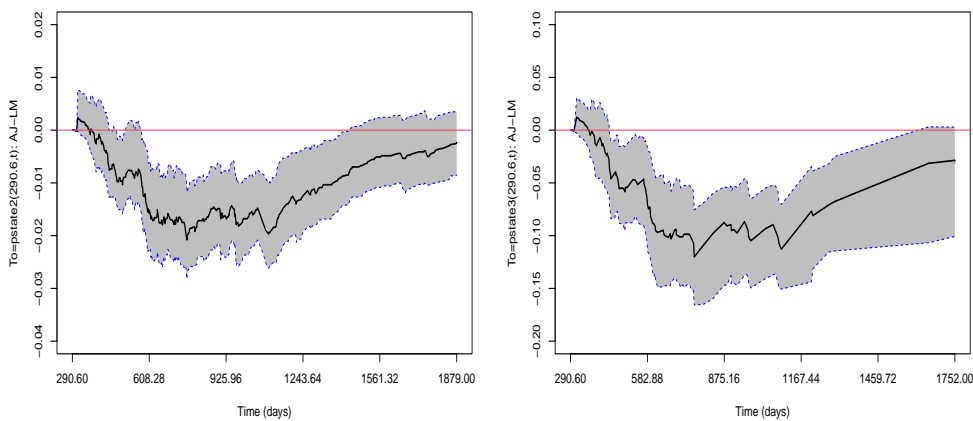


Figure 4: Local graphical test for the Markov condition for colon cancer data, for s equal to 290.6, the third percentile of the sojourn time in the initial state. Transition 1→2 is shown on the left hand side, and transition 2→3 is shown on the right hand side.

the Markovian assumption for the third percentile time (p -value < 0.001). Plots with the graphical local tests for the respective quantiles can be easily obtained using the following input commands:

```
> plot(res4, quantileOrder = 3, to = 2, axis.scale = c(-0.04, 0.02))
#Figure 4 (Left column)

> plot(res5, quantileOrder = 3, to = 3, axis.scale = c(-0.20, 0.10))
#Figure 4 (Right column)
```

The plots shown in Figure 4 display the differences between the AJ and LMAJ estimates for the third quantile (quantileOrder) of the sojourn time in the initial state being in accordance with the p -values obtained in the local tests.

Often, multi-state models include covariates, and it may be the omission of covariate effects that induces apparent non-Markovianity. The methods proposed in this paper can also deal with this problem since discrete covariates can be included in the estimation of the transition probabilities $p_{hj}(s, t)$ by splitting the sample for each level of the covariate and repeating the described procedures for each subsample. As an example, the following input codes show how to obtain the AUC global for individuals with the treatment 'Obs' of covariate 'rx'. To this end, we start to filter the wide data set

| Argument | Description |
|---------------|--|
| data | Multi-state data in msdata format, which is created from the mstate package. It should also contain (dummy coding of) the relevant covariates; no factors allowed. |
| times | Grid of time points at which to compute the statistic. |
| from | The starting state of the transition to check the Markov condition. |
| to | The last state of the considered transition to check the Markov condition. |
| replicas | The number of wild bootstrap replications. This method, proposed by Wu (1986) is commonly applied to solve problems related to heteroscedasticity. Beyersmann and Allignol (2012) found better small sample results using centered Poisson random variables rather than standard normal weights. For further information on the usage of the wild bootstrap, see Titman and Putter (2020). |
| formula | Right-hand side of the formula. If NULL will fit with no covariates (formula="1" will also work), offset terms can also be specified. |
| fn | A list of summary functions to be applied to the individual zbar traces (or a list of lists). |
| fn2 | A list of summary functions to be applied to the overall chi-squared trace. |
| dist | Distribution of wild bootstrap random weights, either poisson for centred Poisson (default), or normal for standard normal. |
| min_time | The minimum time for calculating optimal weights. |
| other_weights | Other (than optimal) weights can be specified here. |

Table 4: Summary of the arguments of the function LR.test.

colonMSM and run the prepMSM function to obtain the subgroups of the new data in the format long. At least, we just need to use the AUC.test function for the specific transitions as follows:

```
> db_wide_obs <- db_wide[db_wide$rx == 'Obs',]
> db_long_obs <- prepMSM(data = db_wide_obs, trans, timesNames, status)
> set.seed(12345)
> res4_obs <- AUC.test(data = db_long_obs, times = 365, from = 1, to = 3,
+                      type='local', replicas = 100, tmat = tmat)
> res4_obs$localTest

      s      1->1      1->2      1->3
1 365  0.6295861  0.0003563306  0.0004307342

> set.seed(12345)
> res5_obs <- AUC.test(data = db_long_obs, times = 365, from = 2, to = 3,
+                      type = 'local', replicas=100, tmat = tmat)
> res5_obs$localTest

      s      2->2      2->3
1 365  5.935795e-05  0.0001894722
```

The **markovMSM** package can also be used to compute the results of the global and local tests proposed by Titman and Putter (2020) which are based on log-rank statistics. A summary of the arguments of the LR.test function is shown in Table 4. The following input commands illustrate the usage of LR.test function to implement these tests:

```
> set.seed(1234)
> res6 <- LR.test(data = db_long, times = 180, from = 2, to = 3, replicas = 1000)
> res6$globalTestLR
[1] 0.047
```

As we can see, the probability value of the global tests based on the log-rank statistics confirms the possible failure of Markovianity for lower values of s ($P=0.047 < 0.05$).

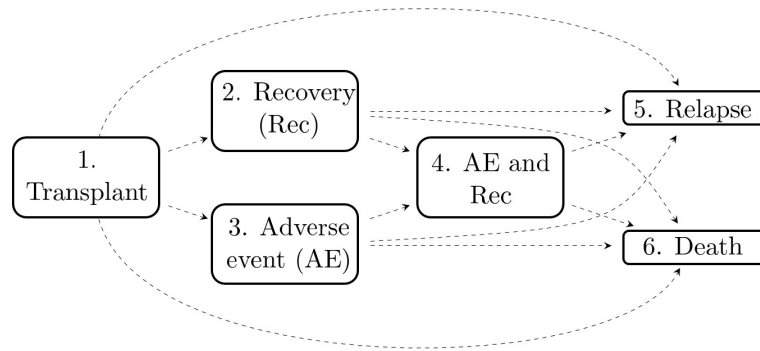


Figure 5: A six-state model given by the Blood and Marrow Transplantation (EBMT) data set for leukemia patients after bone marrow transplantation.

Tests for Markov assumption for more complex multi-state models

In this section, we use two data sets to illustrate the extension of the previous functions to more complex multi-state models. As a first example, we consider the data of 2279 patients transplanted by the European Society for Blood and Marrow Transplantation (EBMT) and, as a second example, data from liver cirrhosis patients subjected to prednisone treatment. Further details on the description of the data can be found in [Putter et al. \(2007\)](#) and [Andersen et al. \(1993\)](#), respectively.

The steps on the usage of the functions are quite similar to those introduced for the illness-death model. To extend the proposed local and global tests to more complex models, we make use of the LMAJ estimator that produces consistent estimates of the transition probabilities in the case of non-Markovianity of the process.

We start to consider the data set comprising 2279 patients who suffered from blood cancers and who were treated at the EBMT between 1985 and 1998 after a transplant. The movement of the patients among the six states can be modeled through the multi-state model with the following six states: 'Alive and in remission, no recovery or adverse event' (state 1); 'Alive in remission, recovered from the treatment' (state 2); 'Alive in remission, occurrence of the adverse event' (state 3); 'Alive, both recovered and adverse event' (state 4); 'Alive, in relapse' (treatment failure) (state 5) and 'Dead (treatment failure)' (state 6). In total, there are 12 transitions: three intermediate events given by recovery (Rec), adverse event (AE) and a combination of the two (AE and Rec), and two absorbing states: relapse and death (Figure 5).

Since the original data ebmt4 is in the wide format, before implementing a global test we need to convert it into the long format using functions `transMatMSM`, `prepMSM` before using function `AUC.test` with the argument `type='global'`:

```

> data("ebmt4")
> db_wide <- ebmt4
> positions <- list(c(2, 3, 5, 6), c(4, 5, 6), c(4, 5, 6),
  c(5, 6), c(), c())
> state.names <- c("Tx", "Rec", "AE", "Rec+AE", "Rel", "Death")
> tmat <- transMatMSM(positions, state.names)
> timesNames <- c(NA, "rec", "ae", "recae", "rel", "srv")
> status <- c(NA, "rec.s", "ae.s", "recae.s", "rel.s", "srv.s")
> trans <- tmat
> db_long <- prepMSM(data = db_wide, trans, timesNames, status)
> db_long[1:10,]

```

Data:

| | id | from | to | trans | Tstart | Tstop | time | status |
|---|----|------|----|-------|--------|-------|------|--------|
| 1 | 1 | 1 | 2 | 1 | 0 | 22 | 22 | 1 |
| 2 | 1 | 1 | 3 | 2 | 0 | 22 | 22 | 0 |
| 3 | 1 | 1 | 5 | 3 | 0 | 22 | 22 | 0 |
| 4 | 1 | 1 | 6 | 4 | 0 | 22 | 22 | 0 |
| 5 | 1 | 2 | 4 | 5 | 22 | 995 | 973 | 0 |
| 6 | 1 | 2 | 5 | 6 | 22 | 995 | 973 | 0 |
| 7 | 1 | 2 | 6 | 7 | 22 | 995 | 973 | 0 |

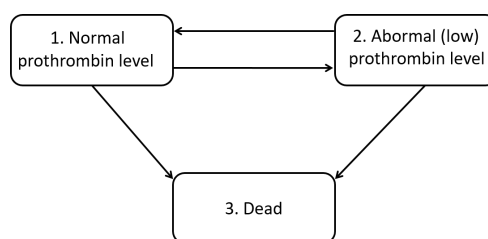


Figure 6: A reversible illness-death model from the data set of patients with liver cirrhosis subjected to prednisone treatment. Reversible transitions are possible in transient states defined by prothrombin levels.

```

8  2  1  2  1  0  12  12  0
9  2  1  3  2  0  12  12  1
10 2  1  5  3  0  12  12  0

```

```

> set.seed(1234)
> res7 <- AUC.test(data = db_long, from = 1, to = 5, type = 'global',
  quantiles = c(0.05, 0.10, 0.20, 0.30, 0.40),
  tmat = tmat, replicas = 100,
  positions = positions, namesStates = state.names,
  timesNames = timesNames, status = status)

```

```

> round(res7$globalTest, 4)
  1->1  1->2  1->3  1->4  1->5
1  0.1417 0.0094 0.0231 0.099 0.0068

```

```

> round(res7$localTests, 4)
  s  1->1  1->2  1->3  1->4  1->5
1  9.9 0.2822 0.6030 0.0176 0.0848 0.0137
2 12.0 0.2833 0.5230 0.0285 0.4643 0.0000
3 15.0 0.0001 0.2106 0.0398 0.0191 0.1193
4 18.0 0.4584 0.0041 0.1763 0.1789 0.0240
5 21.0 0.0060 0.0147 0.4673 0.2363 0.4796

```

The interpretation of the output for the global test and for local tests of the five percentile times is similar to that shown for the illness-death model. Thus, object `res7` is used to get a list with components giving the probability values for the Markov test for all transitions leaving State 1 (i.e., for $1 \rightarrow j$ with $j \in \{1, 2, 3, 4, 5\}$). For instance, the probability value of 0.0094 corresponds to the result of the AUC global test for transition $1 \rightarrow 2$, while 0.0147 is the probability value for the local test, for $s = 21$ for the same transition.

The proposed methods can also be used in reversible multi-state models such as those applied to the data set of liver cirrhosis patients who were included in a randomized clinical trial at several hospitals in Copenhagen between 1962 and 1974. The study aimed to evaluate whether a treatment based on prednisone prolongs survival for patients with cirrhosis [Andersen et al. \(1993\)](#). State 1 corresponds to ‘normal prothrombin level’, State 2 to ‘low (or abnormal) prothrombin level’, and State 3 to ‘dead’. The movement of the patients among these three states can be modeled using the reversible illness-death model shown in Figure 6. The input commands for the local and global tests of the AUC and Log-rank tests are presented as a vignette in the `markovMSM` package.

4. Conclusions

This paper discusses the implementation in R of new methods for checking the Markov assumption in multi-state models proposed by [Soutinho and Meira-Machado \(2021\)](#). Instead of testing, as usual, the Markovianity of the multi-state process by including covariates depending on the history, they propose global and local tests based on the comparison between estimated transition probabilities. This can be done by measuring the discrepancy between the Aalen-Johansen estimator, which gives consistent estimators in Markov processes, and recent approaches that do not rely on this assumption, given by the Landmark and the Landmark Aalen-Johansen estimators.

The use of local tests is recommended whenever the interest is focused on the estimation of the transition probabilities and, in particular, to decide which estimator is the most appropriate to use: the Aalen-Johansen estimator or a robust estimator. A global test, such as the test proposed here, might be preferable for regression purposes. To this end, a common simplifying strategy is to decouple the whole process into various survival models. Then, for all permitted transitions, the transition intensities may be modeled using separate Cox models, assuming the process to be Markovian (also known as the clock forward modeling approach). When the test rejects the Markov assumption, one alternative approach is to use a semi-Markov Cox model in which the future of the process does not depend on the current time but rather on the duration of the current state.

A brief theory underlying the methods has been summarized in this paper, as well as an extensive literature review, which is also indicated. A detailed usage of the three main functions that compose the `markovMSM` package, `AUC.test`, `LR.test` and `PHM.test` are illustrated using three real data sets: `colonIDM`, `ebmt4` and `prothr`. Whereas the first data enables the application of the Markov tests to an illness-death model, the last two cases show how to extend the proposed methods to more complex multi-state models (with more than three states or with reversible transitions). The comparison of the proposed tests to the new methods given by the log-rank test described by Titman and Putter (2020) is also shown. We also present approaches to validate the suspicion of the failure of the Markovianity of the process using graphical tests. The `markovMSM` package is available at the first author's GitHub repository as well as the CRAN repository at <https://cran.r-project.org/web/packages/markovMSM>. Further details on the usage of its functions can be obtained from the corresponding help pages.

Acknowledgments

This research was financed by Portuguese Funds through FCT - "Fundação para a Ciência e a Tecnologia", within Projects projects UIDB/00013/2020, UIDP/00013/2020 and the research grant PD/BD/142887/2018.

Bibliography

- O. Aalen and S. Johansen. An empirical transition matrix for non homogeneous markov and chains based on censored observations. *Scandinavian Journal of Statistics*, 5:141–150, 1978. [p1, 4]
- A. Allignol and A. Latouche. Cran task view: Survival analysis. Version 2022-03-07, URL <http://CRAN.R-project.org/view=Survival>, 2022. [p2]
- P. Andersen and N. Keiding. Multi-state models for event history analysis. *Statistical Methods in Medical Research*, 11(2):91–115, 2002. [p1, 2, 4]
- P. Andersen, S. Esbjerg, and T. Sorensen. Multistate models for bleeding episodes and mortality in liver cirrhosis. *Statistics in Medicine*, 19(4):587–599, 2000. [p2, 4]
- P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer-Verlag, New York, 1993. [p1, 4, 6, 13, 14]
- T. A. Balan and H. Putter. frailtyEM: An R package for estimating semiparametric shared frailty models. *Journal of Statistical Software*, 90(7):1–29, 2019. doi: 10.18637/jss.v090.i07. [p2]
- M. Beyersmann, J. Schumacher and A. Allignol. *Competing Risks and Multistate Models with R*. Springer, New York, 2012. [p12]
- O. Borgan. *Encyclopedia of biostatistics: Aalen-Johansen estimator*. John Wiley & Sons, 2005. [p4]
- S. Chiou, J. Qian, E. Mormino, and R. Betensky. Permutation tests for general dependent truncation. *Computational Statistics & Data Analysis*, 318:308–324, 2018. doi: 10.1016/j.csda.2018.07.012. [p2]
- S. Datta and G. Satten. 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. [p2, 4]
- J. de Uña-Álvarez and L. Meira-Machado. Nonparametric estimation of transition probabilities in the non-markov illness-death model: A comparative study. *Biometrics*, 71(2):364–375, 2015. ISSN 0006-341X. [p2, 4, 5]

- G. L. Hickey, P. Philipson, A. Jorgensen, and R. Kolamunnage-Dona. *joinerML: a joint model and software package for time-to-event and multivariate longitudinal outcomes*, 2018. [p3]
- P. Hougaard. *Analysis of Multivariate Survival Data*. Statistics for Biology and Health. Springer-Verlag, New York, 2000. [p1]
- C. Jackson. Multi-state models for panel data: The msm package for r. *Journal of Statistical Software*, 38:8:1–28, 2011. [p2]
- R. Kay. A markov model for analyzing cancer markers and disease states in survival studies. *Biometrics*, 42(4):457–481, 1986. [p2, 4]
- A. Król and P. Saint-Pierre. SemiMarkov: An R package for parametric estimation in multi-state semi-markov models. *Journal of Statistical Software*, 66(6):1–16, 2015. URL <http://www.jstatsoft.org/v66/i06/>. [p2]
- R. Li and L. Peng. *Handbook of Quantile Regression*. Chapman and Hall/CRC, London, 2017. [p1]
- L. Meira-Machado and M. Sestelo. Estimation in the progressive illness-death model: A nonexhaustive review. *Biometrical Journal*, 61(2):245–263, 2019. doi: 10.1002/bimj.201500038. [p1, 3]
- L. Meira-Machado, J. de Uña-Álvarez, and C. Cadarso-Suárez. Nonparametric estimation of transition probabilities in a non-markov illness-death model. *Lifetime Data Analysis*, 12:325–344, 2006. [p2]
- L. Meira-Machado, J. de Uña-Álvarez, C. Cadarso-Suárez, and P. Andersen. Multi-state models for the analysis of time to event data. *Statistical Methods in Medical Research*, 18:195–222, 2009. [p1]
- C. Moertel, T. Fleming, J. Macdonald, D. Haller, J. Laurie, C. Tangen, J. Ungerleider, W. Emerson, D. Tormey, J. Glick, M. Veeder, and J. Mailliard. Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage iii. colon carcinoma: A final report. *The Annals of Internal Medicine*, 122(5):321–326, 1995. [p6]
- C. G. Moertel, T. R. Fleming, J. S. Macdonald, D. G. Haller, J. A. Laurie, P. J. Goodman, J. S. Ungerleider, W. A. Emerson, D. C. Tormey, J. H. Glick, M. H. Veeder, and J. A. Mailliard. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine*, 322(6):352–358, 1990. [p6]
- D. Nevo and M. Gorfine. Causal inference for semi-competing risks data. *arXiv*, 2020. doi: arXiv:2010.04485. URL <https://doi.org/10.48550/arXiv.2010.04485>. [p1]
- P. Philipson, I. Sousa, P. J. Diggle, P. Williamson, R. Kolamunnage-Dona, R. Henderson, and G. L. Hickey. *joiner: Joint Modelling of Repeated Measurements and Time-to-Event Data*, 2018. URL <https://github.com/graemeleehickey/joiner/>. R package version 1.2.6. [p2]
- H. Putter and C. Spitoni. 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. [p2, 4, 5]
- H. Putter, M. Fiocco, and R. B. Geskus. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, 26(11):2389–2430, 2007. [p1, 6, 13]
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL <http://www.R-project.org/>. [p6]
- D. Rizopoulos. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33, 2010. URL <https://doi.org/10.18637/jss.v035.i09>. [p2]
- M. Rodriguez-Gironde and J. de Uña-Álvarez. A nonparametric test for markovianity in the illness-death model. *Statistics in Medicine*, 31(30):4416–4427, 2012. doi: 10.1002/sim.5619. [p2]
- M. Rodriguez-Gironde and J. de Uña-Álvarez. Methods for testing the markov condition in the illness-death model: a comparative study. *Statistics in Medicine*, 35(20):3549–3562, 2016. doi: 10.1002/sim.6940. [p2]
- G. Soutinho and L. Meira-Machado. Methods for checking the markov condition in multi-state survival data. *Computational Statistics*, 2021. doi: 10.1007/s00180-021-01139-7. [p2, 3, 4, 5, 14]
- G. Soutinho, M. Sestelo, and L. Meira-Machado. survdm: An R package for Inference and Prediction in an Illness-Death Model. *The R Journal*, 13(2):70–89, 2021. doi: 10.32614/RJ-2021-070. URL <https://doi.org/10.32614/RJ-2021-070>. [p2]

- A. Titman and H. Putter. General tests of the markov property in multi-state models. *Bio-statistics*, 2020. doi: 10.1093/biostatistics/kxaa030. [p2, 4, 5, 6, 12, 15]
- H. C. Van Houwelingen. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, 34(1):70–85, 2007. [p4]
- C. Wu. Jackknife, bootstrap, and other resampling methods in regression analysis. *Annals of Statistics*, 14(4):1261–1295, 1986. [p12]
- J. Xu, J. Kalbfleisch, and B. Tai. Statistical analysis of illness-death processes and semicompeting risks data. *Biometrics*, 66:716–727, 2010. doi: 10.1111/j.1541-0420.2009.01340.x. URL <https://doi.org/10.1111/j.1541-0420.2009.01340.x>. [p1]

Gustavo Soutinho
EPIUnit
Institute of Public Health of the University of Porto (ISPUP)
Rua das Taipas 135, 4050-600 Porto
Portugal
ORCID: 0000-0002-0559-1327
gdsoutinho@gmail.com

Luís Meira-Machado
Department of Mathematics & Centre of Mathematics
University of Minho
Campus de Azurém - 4800-058 Guimarães
Portugal
ORCID: 0000-0002-8577-7665
lmachado@math.uminho.pt