Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen-Johansen estimator

Putter, H.; Spitoni, C.

Journal Title
XX(X):2–20

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Non-parametric estimation of transition probabilities in non-Markov multi-state models: the landmark Aalen-Johansen estimator

Hein Putter[1] and Cristian Spitoni[2]

Abstract

The topic non-parametric estimation of transition probablities in non-Markov multi-state models has seen a remarkable surge of activity recently. Two recent papers have used the idea of subsampling in this context. The first paper, by de Uña Álvarez and Meira-Machado, uses a procedure 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. The second, by Titman, derived estimators of transition probabilities that are consistent in general non-Markov multi-state models. Here we show that the same idea of subsampling, used in both these papers, combined with the Aalen-Johansen estimate of the state occupation probabilities derived from that subset, can also be used to obtain a relatively simple and intuitive procedure which we term landmark Aalen-Johansen (LMAJ). We show that the LMAJ estimator yields a consistent estimator of the transition probabilities in general non-Markov multi-state models under the same conditions as needed for consistency of the Aalen-Johansen estimator of the state occupation probabilities. Simulation studies show that the LMAJ estimator has good small sample properties and is slightly more efficient than the other estimators.

Keywords

Multi-state model, transition probability, Markov assumption

Prepared using sagej.cls *[Version: 2015/06/09 v1.01]*

Introduction

Multi-state models are finding increased application in medical research. They allow a detailed view of the disease or recovery process of a patient, and they can be used to obtain prediction probabilities of future events, after a given event history. A number of reviews on multi-state models are available in the literature[1–5]. The relevant quantities for these prediction probabilities in multi-state terminology are the transition probabilities, the probabilities to be in a state m at time t, given that the patient is in state ℓ at an earlier time s. When the multi-state model is Markov, an elegant theory connects the transition intensities of the multi-state model to the transition probabilities, leading to the Aalen-Johansen estimator[6].

When the multi-state model is Markov, the Aalen-Johansen estimator gives consistent estimators of the transition probabilities. When the multi-state model is non-Markov, this is no longer the case. Meira-Machado et al.[7] considered estimation of the transition probabilities for a non-Markov irreversible illness-death model. Their procedure was based on expressing the transition probabilities of interest in terms of expectations of transformations of the joint distribution of the time to absorption and the sojourn time in the initial state, and replacing these expressions by weighted averages. They showed superior performance of their non-Markov estimators over the Aalen-Johansen estimator in case of strong violation of the Markov assumption. Allignol et al.[8] defined a competing risks process (which is by its nature always Markov) for which the cumulative incidences relate in a certain way to the transition probabilities of interest. Both methods require that the support of the censoring distribution is contained in the support of the lifetime distribution, an assumption that is unlikely to hold in most medical applications, because of limited follow-up of patients. Two recent papers have improved on these results by removing the restrictive support assumption. The paper by de Uña Álvarez and Meira-Machado[9] considers an irreversible illness-death model and proposes – among others – a subsampling approach where a selection is made of the data consisting of subjects occupying a given state at a particular time; based on this subset an estimator proposed by Pepe[10] consisting of a difference between two Kaplan-Meier estimates is proposed for one of the transition probabilities. Titman[11] extended and improved on[8]

[1]Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands
[2]Department of Mathematics, University of Utrecht, Budapestlaan 6, 3584 CD Utrecht, The Netherlands

Corresponding author:
Hein Putter, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands
Email: h.putter@lumc.nl

Prepared using sagej.cls

by also allowing extension to general multi-state models. Although not explicitly mentioned, Titman's estimators are also based on subsampling.

Although the non-parametric Aalen-Johansen estimator[6] will not in general give consistent estimators of the transition probabilities in non-Markov multi-state models, Datta and Satten[12] have shown that, even for non-Markov multi-state models, the estimator of state occupation probabilities derived from the non-parametric Aalen-Johansen estimator is consistent. The paper by Glidden[13] provides further understanding of this result and presents asymptotic results and estimators of pointwise standard errors and simultaneous confidence bands. In this paper we show that a relatively simple and intuitive procedure that we call landmark Aalen-Johansen (LMAJ) will also provide consistent estimators of transition probabilities for general multi-state models. As in[9] and[11], the procedure is based on subsampling; it selects subjects fulfilling the requirements of being in a given state (or set of states). Within this subset, estimates of the state occupation probabilities are obtained using the Aalen-Johansen estimator. Since the idea of selecting subjects in a given state at a given (landmark) time is akin to landmarking[14–16], we refer to the new estimator as the landmark Aalen-Johansen (LMAJ) estimator. The LMAJ estimator makes no assumptions on the support of the censoring distribution and is defined for arbitrary multi-state models. We show how the LMAJ estimator compares with the other estimators[9;11] by including the LMAJ estimator in the same simulation set-up as Titman[11], and two additional scenarios, and we apply the LMAJ estimator in data from a randomized clinical trial in breast cancer. For easy comparison of the LMAJ estimator with the aforementioned estimators, we refer to them with the same abbreviations as used by Titman[11]: CP (conditional Pepe) for the subsampling estimator of de Uña Álvarez and Meira-Machado[9] and NM (non-Markov) for the estimator of Titman[11].

The landmark Aalen-Johansen estimator

We broadly follow notation of Glidden[13] and define $\tilde{X}(t)$ to be a random multi-state process, taking values in the state space $1, \ldots, K$, with K finite. The multi-state process has right-continuous paths, and a finite number of transitions. For $i = 1, \ldots, n$, and $j, k = 1, \ldots, K$, with $j \neq k$, the counting processes

$$\tilde{N}_{ijk}(t) = \#\{u \leq t, \tilde{X}_i(u-) = j, \tilde{X}_i(u) = k\}$$

count the number of direct transitions of subject i from state j to state k up to and including time t, and

$$\tilde{Y}_{ij}(t) = 1\{\tilde{X}_i(t-) = j\}$$

is the at-risk process of subject i corresponding to state j. The sigma-algebra generated by the counting and at-risk processes defines the filtration

$$\mathfrak{F}_t = \sigma\{\tilde{N}_{ijk}(u), \tilde{Y}_{ij}(u), 0 \leq u \leq t, i = 1, \ldots, n, j, k = 1, \ldots, K, j \neq k\}.$$

The transition hazards are defined by

$$\tilde{\lambda}_{jk}(t \,|\, \mathfrak{F}_{t-}) = \lim_{\Delta t \downarrow 0} P(\tilde{X}(t + \Delta t) = k \,|\, \tilde{X}(t) = j, \mathfrak{F}_{t-})/\Delta t. \tag{1}$$

In general (non-Markov) multi-state models, these transition hazards will depend not only on the present state j at time t, but also on the further past $\mathfrak{F}_{t-}$. When the multi-state process is Markov, (1) simplifies to $\tilde{\lambda}_{jk}(t) = \lim_{\Delta t \downarrow 0} P(\tilde{X}(t + \Delta t) = k \,|\, \tilde{X}(t) = j)/\Delta t$. The transition probabilities are defined by

$$P_{\ell m}(s, t \,|\, \mathfrak{F}_{s-}) = P(\tilde{X}(t) = m \,|\, \tilde{X}(s) = \ell, \mathfrak{F}_{s-}).$$

For Markov models, the transition probabilities simplify to $P_{\ell m}(s, t) = P(\tilde{X}(t) = m \,|\, \tilde{X}(s) = \ell)$.

When the multi-state model is not Markov, the transition probability $P_{\ell m}(s, t \,|\, \mathfrak{F}_{s-})$ will depend on the past before time s, $\mathfrak{F}_{s-}$. For instance, when the process is a Markov renewal process, $P_{\ell m}(s, t \,|\, \mathfrak{F}_{s-})$ will crucially depend on the time at which state ℓ was entered before time s, because that will determine the duration in state ℓ. We would like to emphasize at this point that in such a case one should always try to take into account the extra relevant information (here the time of entry into state ℓ), both in the target of inference and in the estimation procedure; for instance in Markov renewal processes explicit expressions for estimators of transition probabilities are also available[17]. Even when the multi-state model is not Markov, however, the transition probability $P_{\ell m}(s, t)$ may be relevant as a summary of different transition probabilities $P_{\ell m}(s, t \,|\, \mathfrak{F}_{s-})$, in case 1) the extra relevant information in $\mathfrak{F}_{s-}$ is not available, or 2) when interest is in an average over the histories $\mathfrak{F}_{s-}$ of $P_{\ell m}(s, t \,|\, \mathfrak{F}_{s-})$, or 3) when it is unknown how $\tilde{\lambda}_{jk}(t \,|\, \mathfrak{F}_{t-})$ depends on the history; in practice it will often be uncertain whether or not the multi-state model is Markov and in such cases an estimator that is robust against possible non-Markovianity would be useful.

The sentence "average over the histories $\mathfrak{F}_{s-}$ of $P_{\ell m}(s, t \,|\, \mathfrak{F}_{s-})$" is admittedly not very precise. In particular instances, when the nature of violation of the Markov assumption is known, the statement can be made precise. To give an example of what is meant, consider the irreversible illness-death (Markov renewal) model, where the transition $\tilde{\lambda}_{23}(t \,|\, \mathfrak{F}_{t-})$ from the illness to the death state depends (only) on the duration in state 2. Then $P_{23}(s, t \,|\, \mathfrak{F}_{s-})$ depends on $\mathfrak{F}_{s-}$ only through the time of entry, say T_2, in state

2. In this instance, we have

$$P_{23}(s,t) = \int_0^s P_{23}(s,t \,|\, T_2 = t_2) f(t_2 \,|\, X(s) = 2) dt_2,$$

where $f(t_2 \,|\, X(s) = 2)$ represents the density of T_2, given the subject is in state 2 at time $s > t_2$. Other types of violations of the Markov assumption will call for different dependencies of $P_{\ell m}(s,t \,|\, \mathfrak{F}_{s-})$ on $\mathfrak{F}_{s-}$, and hence different kind of averages.

For a Markov model, define the cumulative transition hazards $\tilde{\Lambda}_{jk}(t) = \int_0^t \tilde{\lambda}_{jk}(u)du$ and gather all of them in the $K \times K$ matrix $\tilde{\mathbf{\Lambda}}(t)$ with (j,k)th off-diagonal element $\tilde{\Lambda}_{jk}(t)$ and (j,j)th diagonal element $\tilde{\Lambda}_{jj}(t) = -\sum_{k \neq j} \tilde{\Lambda}_{jk}(t)$. Similarly define the $K \times K$ matrix $\mathbf{P}(s,t)$ with (ℓ,m)th element $P_{\ell m}(s,t)$. Then[6] the matrix of transition probabilities can be written as a matrix product integral of the transition hazards, as

$$\mathbf{P}(s,t) = \prod_{s < u \leq t} \left(\mathbf{I} + d\tilde{\mathbf{\Lambda}}(u) \right).$$

The vector $\mathbf{P}(t)$ of state occupation probabilities, with mth element $P_m(t) = P(\tilde{X}(t) = m)$, can be retrieved through $\mathbf{P}(t) = \pi(0)\mathbf{P}(0,t)$, with $\pi(0)$ a $1 \times K$ vector with kth element $\pi_k(0) = P(\tilde{X}(0) = k)$. Together, we have the relation

$$\mathbf{P}(t) = \pi(0) \prod_{0 < u \leq t} \left(\mathbf{I} + d\tilde{\mathbf{\Lambda}}(u) \right). \tag{2}$$

Datta & Satten[12] showed that also for non-Markov multi-state processes, the state occupation probability vector follows a relation like (2), but with $\mathbf{\Lambda}(\cdot)$ replaced by the partly conditional transition rates $\mathbf{\Lambda}(\cdot)$[18], where

$$\lambda_{jk}(t) = \lim_{\Delta t \downarrow 0} P(\tilde{X}(t + \Delta t) = k \,|\, \tilde{X}(t) = j)/\Delta t, \;\; \Lambda_{jk}(t) = \int_0^t \lambda_{jk}(u)du,$$

for the transition rate from state j to state k. In contrast to the transition rates in (1) the partly conditional transition rates condition only on the current state, not on the further history $\mathfrak{F}_{t-}$, and can be thought of as complex weighted averages of transition hazards over all possible histories.

The observed data consist of right censored versions of the multi-state process. Let C_i be a right censoring time, for $i = 1, \ldots, n$, assume that $\tilde{X}_i(\cdot)$ and C_i are independent and identically distributed, and define $H_i(t) = 1\{C_i \geq t\}$ and the censored multi-state, counting and at-risk processes $X_i(t) =$

$\tilde{X}(t \wedge C_i)$,

$$\begin{aligned} N_{ijk}(t) &= \#\{u \le t, X_i(u-) = j, X_i(u) = k, H_i(u) = 1\}, \\ Y_{ij}(t) &= 1\{X_i(t-) = j, H_i(t) = 1\}. \end{aligned}$$

To define Aalen-Johansen type estimators, it is convenient to gather the counting and at-risk processes into $K \times K$ matrices. Define $\mathbf{N}_i(t)$ to be the matrix with off-diagonal (j,k)th element $N_{ijk}(t)$ and (j,j)th diagonal element $N_{ijj}(t) = -\sum_{k \neq j} N_{ijk}(t)$, and define $\mathbf{Y}_{iD}(t)$ to be the diagonal matrix with (j,j)th diagonal element equal to $Y_{ij}(t)$. Estimators of the cumulative partly conditional transition rates are obtained by defining $\overline{N}_{jk}(t) = \sum_{i=1}^n N_{ijk}(t)$, $\overline{Y}_j(t) = \sum_{i=1}^n Y_{ij}(t)$, their matrix versions $\overline{\mathbf{N}}(t)$ and $\overline{\mathbf{Y}}_D(t)$, and finally

$$\widehat{\mathbf{\Lambda}}(t) = \int_0^t \overline{\mathbf{Y}}_D^{-1}(u) d\overline{\mathbf{N}}(u).$$

With $\hat{\pi}(0)$ the $1 \times K$ vector containing the empirical proportions $\hat{\pi}_k(0) = n^{-1} \sum_{i=1}^n 1\{\tilde{X}_i(0) = k\}$, under appropriate conditions the empirical counterpart of (2),

$$\widehat{\mathbf{P}}(t) = \hat{\pi}(0) \prod_{0 < u \le t} \left(\mathbf{I} + \Delta \widehat{\mathbf{\Lambda}}(u) \right) \tag{3}$$

provides a consistent estimator of the state occupation probabilities $P(\tilde{X}(t) = m)$ [12].

We are ready to formulate our landmark Aalen-Johansen estimator of the transition probabilities $P_{\ell m}(s,t) = P(\tilde{X}(t) = m \,|\, \tilde{X}(s) = \ell)$. For fixed s and ℓ, the estimator is based on re-estimated partly conditional rates obtained from selecting subjects with $X_i(s) = \ell$. We will use the superscript (LM) to denote versions of counting and at risk processes and of estimators based on a landmark data set which selects subjects who are at state ℓ at time s, suppressing in the notation that this selection depends on the fixed ℓ and s. Thus, the landmark based versions of $\overline{N}_{jk}(t)$, $\overline{Y}_j(t)$ and $\widehat{\mathbf{\Lambda}}(t)$ are defined as

$$\begin{aligned} \overline{N}_{jk}^{(\mathrm{LM})}(t) &= \sum_{i=1}^n N_{ijk}(t) 1\{X_i(s) = \ell\}, \\ \overline{Y}_j^{(\mathrm{LM})}(t) &= \sum_{i=1}^n Y_{ij}(t) 1\{X_i(s) = \ell\}, \\ \widehat{\mathbf{\Lambda}}^{(\mathrm{LM})}(t) &= \int_0^t {\overline{\mathbf{Y}}_D^{(\mathrm{LM})}}^{-1}(u) d\overline{\mathbf{N}}^{(\mathrm{LM})}(u), \end{aligned} \tag{4}$$

where $\overline{\mathbf{N}}^{(\mathrm{LM})}(t)$ and $\overline{\mathbf{Y}}_D^{(\mathrm{LM})}(t)$ are matrices containing as elements $\overline{N}_{jk}^{(\mathrm{LM})}(t)$ and $\overline{Y}_j^{(\mathrm{LM})}(t)$, arranged as in $\overline{\mathbf{N}}(t)$ and $\overline{\mathbf{Y}}_D(t)$. Finally, the LMAJ estimator is given by

$$\widehat{P}_{\ell m}^{\mathrm{LMAJ}}(s,t) = \hat{\pi}^{(\mathrm{LM})}(s) \prod_{s<u\leq t} \left(\mathbf{I} + \Delta\widehat{\mathbf{\Lambda}}^{(\mathrm{LM})}(u)\right), \tag{5}$$

with $\hat{\pi}^{(\mathrm{LM})}(s)$ a $1 \times K$ vector with $\hat{\pi}_\ell^{(\mathrm{LM})}(s) = 1$, and other values equal to 0.

In the appendix we prove consistency of $\widehat{P}_{\ell m}^{\mathrm{LMAJ}}(s,t)$, under the same assumptions as needed for consistency of the Aalen-Johansen estimator of the state occupation probabilities [12], plus the additional assumption that $P(\tilde{X}(s) = \ell) > 0$.

Standard errors

Glidden [13] argues that the Greenwood estimators of the pointwise standard errors of the Aalen-Johansen estimator of the state occupation probabilities remain valid also if the Markov assumption is violated. We claim that the same is true for the LMAJ estimator. The simulation studies, reported below, corroborate this claim. For simultaneous confidence bands more elaborate methods need to be used [13].

Generalized conditional probabilities

Titman [11] considers more generally estimators of $P_{\mathcal{L}\mathcal{M}}(s,t) = P(\tilde{X}(t) \in \mathcal{M} \mid \tilde{X}(s) \in \mathcal{L})$. It is not difficult to see that the proof of consistency of the LMAJ estimator (see Appendix) can be extended in a straightforward way to yield consistency of $\widehat{P}_{\mathcal{L}m}^{\mathrm{LMAJ}}(s,t)$ obtained by replacing $1\{X_i(s) = \ell\}$ by $1\{X_i(s) \in \mathcal{L}\}$ in the definitions of (4). Also, $\hat{\pi}^{(\mathrm{LM})}(s)$ in (5) should be replaced by the vector of relative proportions of subjects in the states $\ell \in \mathcal{L}$ at time s. Finally, a consistent estimator of $P_{\mathcal{L}\mathcal{M}}(s,t)$ follows by defining $\widehat{P}_{\mathcal{L}\mathcal{M}}^{\mathrm{LMAJ}}(s,t) = \sum_m \widehat{P}_{\mathcal{L}m}^{\mathrm{LMAJ}}(s,t)$.

Comparison with CP and NM for the irreversible illness-death model

If there is no censoring between s and t, both the CP (for irreversible illness-death models) and the NM and LMAJ estimators of $P_{\ell m}(s,t)$ (for general multi-state models) reduce to the proportion (among those in state ℓ at time s) in state m at time t. In fact this holds more generally for $\widehat{P}_{\mathcal{L}\mathcal{M}}^{\mathrm{NM}}(s,t)$ and $\widehat{P}_{\mathcal{L}\mathcal{M}}^{\mathrm{LMAJ}}(s,t)$. In the presence of censoring the three estimators may differ.

It is instructive to contrast the CP, NM and LMAJ estimators, in the case of an irreversible illness-death model. It is not difficult to see that all three methods give identical estimates for $P_{22}(s,t)$ and $P_{23}(s,t) = 1 - P_{22}(s,t)$, so we concentrate on estimates of $P_{11}(s,t)$, $P_{12}(s,t)$ and $P_{13}(s,t)$. We consider a small example data set, shown in Table 1, and take $s = 1.5$. The table shows the times t_2

id	t_2	t_3
1	2	5
2	3	4+
3	–	7
4	6	8
5	1	9

Table 1. A small data set used for illustration.

and t_3 at which states 2 and 3 were entered. The $4+$ for t_3 of subject 2 means that the subject was censored (in state 2) at time $t = 4$; the – for t_2 of subject 3 means that state 2 was never reached because the subject went directly to state 3.

All estimators only consider subjects 1–4, since subject 5 is in state 2 at time $s = 1.5$. Recall the definitions of $\overline{N}_{jk}^{(\mathrm{LM})}(t)$ and $\overline{Y}_{j}^{(\mathrm{LM})}(t)$ from Equation (4), taking $\ell = 1$ and $s = 1.5$ to define the landmark data set, and in addition, define

$$\begin{aligned} N_{i\mathcal{J}k}(t) &= \#\{u \le t, X_i(u-) \in \mathcal{J}, X_i(u) = k, H_i(u) = 1\}, \\ Y_{i\mathcal{J}}(t) &= 1\{X_i(t-) \in \mathcal{J}, H_i(t) = 1\}, \end{aligned}$$

and

$$\begin{aligned} \overline{N}_{\mathcal{J}k}^{(\mathrm{LM})}(t) &= \sum_{i=1}^{n} N_{i\mathcal{J}k}(t) 1\{X_i(s) = \ell\}, \\ \overline{Y}_{\mathcal{J}}^{(\mathrm{LM})}(t) &= \sum_{i=1}^{n} Y_{i\mathcal{J}}(t) 1\{X_i(s) = \ell\}. \end{aligned}$$

Then, with $\overline{N}_{1\bullet}^{(\mathrm{LM})}(t) = \overline{N}_{12}^{(\mathrm{LM})}(t) + \overline{N}_{13}^{(\mathrm{LM})}(t)$, we can define the conditional Kaplan-Meier survival functions

$$\begin{aligned} \widehat{S}_1(t \,|\, s) &= \prod_{s<u\le t} \Big(1 - \frac{d\overline{N}_{1\bullet}^{(\mathrm{LM})}(u)}{\overline{Y}_{1}^{(\mathrm{LM})}(u)}\Big), \\ \widehat{S}_{\{12\}}(t \,|\, s) &= \prod_{s<u\le t} \Big(1 - \frac{d\overline{N}_{\{12\}}^{(\mathrm{LM})}(u)}{\overline{Y}_{\{12\}}^{(\mathrm{LM})}(u)}\Big), \\ \widehat{S}_2(t \,|\, s) &= \prod_{s<u\le t} \Big(1 - \frac{d\overline{N}_{23}^{(\mathrm{LM})}(u)}{\overline{Y}_{2}^{(\mathrm{LM})}(u)}\Big). \end{aligned}$$

The first estimates the conditional probability of remaining in state 1, the second of remaining in state 1 or 2, the third of remaining in state 2. With these definitions, we can see that estimators for $P_{11}(s,t)$ and $P_{13}(s,t)$ are the same for CP and NM, given by

$$\widehat{P}_{11}^{\text{CP}}(s,t) = \widehat{P}_{11}^{\text{NM}}(s,t) = \widehat{S}_1(t \,|\, s),$$

and

$$\widehat{P}_{13}^{\text{CP}}(s,t) = \widehat{P}_{13}^{\text{NM}}(s,t) = 1 - \widehat{S}_{\{12\}}(t \,|\, s).$$

Estimators for $P_{12}(s,t)$ differ between CP and NM; for CP we simply have $\widehat{P}_{12}^{\text{CP}}(s,t) = 1 - \widehat{P}_{11}^{\text{CP}}(s,t) - \widehat{P}_{13}^{\text{CP}}(s,t)$, while the definition of NM gives

$$\widehat{P}_{12}^{\text{NM}}(s,t) = \widehat{S}_{\{12\}}(t \,|\, s)\hat{\pi}_{2\,|\,1}(t \,|\, s),$$

where $\hat{\pi}_{2\,|\,1}(t \,|\, s)$ is the proportion of subjects, among those not yet absorbed in state 3 and under follow-up at time t and having started in state 1 at time s, who are in state 2 at time t. Note that, in contrast with CP, NM is not guaranteed to satisfy $\widehat{P}_{11}^{\text{NM}}(s,t) + \widehat{P}_{12}^{\text{NM}}(s,t) + \widehat{P}_{13}^{\text{NM}}(s,t) = 1$. Indeed, for $t = 6$, we have $\widehat{P}_{11}^{\text{NM}}(s,t) = 0.250$, $\widehat{P}_{12}^{\text{NM}}(s,t) = 0.333$ and $\widehat{P}_{13}^{\text{NM}}(s,t) = 0.333$, which sums up to less than 1. The CP estimator has $\widehat{P}_{12}^{\text{CP}}(s,t) = 0.417$. The LMAJ estimator has the same estimate of $P_{11}(s,t)$, namely $\widehat{P}_{11}^{\text{LMAJ}}(s,t) = \widehat{S}_1(t \,|\, s)$, and the more complicated estimators

$$\begin{aligned}
\widehat{P}_{12}^{\text{LMAJ}}(s,t) &= \sum_{s<u\leq t} \frac{d\overline{N}_{12}^{(\text{LM})}(u)}{\overline{Y}_1^{(\text{LM})}(u)} \widehat{S}_1(u- \,|\, s)\widehat{S}_2(t \,|\, u), \\
\widehat{P}_{13}^{\text{LMAJ}}(s,t) &= \sum_{s<u\leq t} \frac{d\overline{N}_{12}^{(\text{LM})}(u)}{\overline{Y}_1^{(\text{LM})}(u)} \widehat{S}_1(u- \,|\, s)\left(1 - \widehat{S}_2(t \,|\, u)\right) \\
&\qquad + \sum_{s<u\leq t} \frac{d\overline{N}_{13}^{(\text{LM})}(u)}{\overline{Y}_1^{(\text{LM})}(u)} \widehat{S}_1(u- \,|\, s).
\end{aligned}$$

Estimates of $\widehat{P}_{12}^{\text{LMAJ}}(s,t)$ and $\widehat{P}_{13}^{\text{LMAJ}}(s,t)$ at $t = 5$ for our example data set are given by 0.25 and 0.50, respectively.

It is also of interest to see at which time points the different estimators can change value. The most notable differences can again be seen for the estimates of $P_{12}(s,t)$. Considering only subjects who are in state 1 at time s (for all three estimators), we see that $\widehat{P}_{12}^{\text{LMAJ}}(s,t)$ changes value only at time points t at which a $1 \to 2$ transition is observed. In contrast, both $\widehat{P}_{12}^{\text{CP}}(s,t)$ and $\widehat{P}_{12}^{\text{NM}}(s,t)$ can change value at all time points t at any transition time point, be it a $1 \to 2$, $1 \to 3$, or $2 \to 3$ transition.

The computations from this small data set illustrate that $\widehat{P}_{11}^{\text{NM}}(s,t) + \widehat{P}_{12}^{\text{NM}}(s,t) + \widehat{P}_{13}^{\text{NM}}(s,t) = 1$ is not guaranteed for Titman's estimator. This unfavorable property originates from the construction of these probabilities, which uses a possibly different competing risk process (depending on time s and the state ℓ that is occupied at time s, but also on the target state m) for each transition probability $P_{\ell m}(s,t)$ of interest. From our simulation studies, reported in the next section, it became clear that replications for which $\sum_m \widehat{P}_{\ell m}^{\text{NM}}(s,t) \neq 1$ were common (both smaller and larger than 1), but deviations from 1 were usually very small. If interest is in one particular transition probability $P_{\ell m}(s,t)$, not in $P_{\ell m}(s,t)$ for all states m, the fact that transition probabilities do not add up to one should not be a real problem in practice.

Simulation results

Three sets of simulations were performed, comparing the landmark Aalen-Johansen estimator with the CP and NM estimators.

Irreversible illness-death model

The objective of the first simulation study was to replicate the first simulation study of Titman[11] and to add the new landmark Aalen-Johansen estimator. The set-up is exactly as in Titman[11]. Briefly, data were simulated from a three-state irreversible illness-death model (states numbered here as 1=healthy, 2=illness, 3=death). The same three processes, termed here Markov, Frailty, and non-Markov, were considered. The Markov process was based on a time-homogeneous process with intensities $\alpha_{12} = 0.12$, $\alpha_{13} = 0.03$ and $\alpha_{23} = 0.1$. For the frailty model, all three intensities were multiplied by a common gamma frailty Z with unit mean and variance 2. The other non-Markov process also has the same intensities as the Markov process, except that $\tilde{\lambda}_{13}(t)$ depends on the state occupied at time 4; $\tilde{\lambda}_{13}(t) = 0.05$ if $\tilde{X}(4) = 0$, and 0.1 otherwise. Two different independent right-censoring mechanisms were applied: Unif, where right-censoring times were uniform on $(5, 40)$, and Exp, where right-censoring times were exponential with rate 0.04. Each scenario used sample sizes $n = 200$ and $n = 500$, all starting in state 1. Table 2 reports bias, root mean squared error (RMSE) and empirical coverage of nominal 95% confidence intervals (all $\times 100$) across $M = 5000$ replicated data sets for four methods of estimating $P_{12}(\tau_{0.15}, \tau_{0.45})$, where $\tau_{0.15}$ and $\tau_{0.45}$ are the 15th and 45th percentile, respectively, of the time-to-absorption (state 3) distribution. Values of $\tau_{0.15}$ and $\tau_{0.45}$ were taken from the supplementary material[11]. The four methods considered are the Aalen-Johansen estimator (AJ), the new subsampling method of de Uña Àlvarez & Meira-Machado[9] (CP), the method of Titman[11] (NM) and the new proposed landmark Aalen-Johansen method (LMAJ). The simulations also included the estimator proposed by Allignol et al.[8], but results were not included here, because they clearly underperformed. The same conclusions as

				AJ			CP			NM			LMAJ		
Model	Truth	Cens	n	Bias	RMSE	Cov	Bias	RMSE	Cov	Bias	RMSE	Cov	Bias	RMSE	Cov
Markov	0.3497	Unif	200	-0.086	4.166	94.2	-0.093	4.727	94.2	0.046	4.841	94.9	-0.061	4.739	94.8
			500	0.072	2.691	94.6	0.070	3.051	94.6	0.111	3.105	94.8	0.081	3.045	94.8
		Exp	200	-0.031	4.702	94.3	-0.074	5.395	93.8	0.146	5.545	94.2	-0.019	5.373	94.6
			500	0.031	2.958	95.0	-0.051	3.338	95.1	0.042	3.463	94.9	-0.033	3.345	95.3
Frailty	0.1722	Unif	200	-0.448	3.012	93.0	-0.019	3.300	94.4	0.048	3.444	94.3	0.004	3.291	95.0
			500	-0.416	1.949	93.5	0.017	2.091	95.0	0.034	2.171	95.1	0.029	2.088	95.1
		Exp	200	-0.413	3.427	92.7	-0.032	3.793	93.6	0.062	3.960	93.4	0.015	3.784	94.2
			500	-0.448	2.224	92.4	-0.038	2.412	93.4	0.008	2.544	93.7	-0.018	2.406	93.8
non-Markov	0.3566	Unif	200	4.810	6.578	83.4	0.013	5.225	94.1	0.222	5.339	94.5	0.027	5.211	95.2
			500	4.822	5.577	63.0	-0.052	3.921	94.5	0.023	3.383	94.6	-0.046	3.292	94.9
		Exp	200	4.699	6.907	86.8	0.033	5.937	93.6	0.282	6.124	94.2	0.036	5.915	94.4
			500	4.868	5.807	69.8	0.049	3.916	94.3	0.142	3.981	95.0	-0.031	3.887	95.0

Table 2. Bias, root mean squared error (RMSE) and coverage (all $\times 100$) of the Aalen-Johansen (AJ), the methods by de Uña Àlvarez & Meira-Machado[9] (CP) and Titman[11] (NM) and the new proposed landmark Aalen-Johansen method (LMAJ) for estimating $P_{12}(\tau_{0.15}, \tau_{0.45})$. Truth refers to the true value of $P_{12}(\tau_{0.15}, \tau_{0.45})$.

in Titman[11] apply. In addition, the landmark Aalen-Johansen estimator performs similarly to CP and slightly outperforms NM.

Reversible Markov renewal illness-death model

The second set of simulations was based on a Markov renewal process, with a reversible illness-death model, containing an additional recovery transition from illness (state 2) to healthy (state 1), compared to the illness-death model in Section 3. The hazard rates $\tilde{\lambda}_{jk}(t,d)$, with t time from start, and d duration in the state (sojourn time), were chosen as

$$\tilde{\lambda}_{jk}(t,d) = \alpha_{jk}\beta_j \exp(-\alpha_{jk} d^{\beta_j - 1}), \;\; j = 1,2, k = 1,2,3, j \neq k,$$

i.e., a Weibull hazard with duration d as time scale and no dependence on t. The shape parameters were chosen to be identical for both transitions from state 1 (β_1) and for both transitions from state 2 (β_2). When $\beta_1 = \beta_2 = 1$, hazards are exponential and the model is Markov. For $\beta_1 = \beta_2 = 1$, we chose $\alpha_{12} = 0.12, \alpha_{13} = 0.03, \alpha_{23} = 0.09$ and $\alpha_{21} = 0.06$. Values chosen for β_1 and β_2 were 1, 1.5 and 1, 0.5, respectively. For β_1 and β_2 different from 1, the α_{jk}'s were adjusted so that the expected sojourn times in states 1 and 2 remained the same and the ratios between α_{12} ad α_{13} and between α_{21} and α_{23} also remained the same. Data of $n = 200$ and $n = 500$ subjects were generated, all starting from state 1. Censoring was independent and uniform on $(5, 40)$.

Table 3 shows bias, RMSE and coverage (all $\times 100$) across 5000 replications of $P_{\ell m}(s,t)$ for $s = 5$ and $t = 15$, comparing the Aalen-Johansen estimator (AJ), the non-Markov estimator of Titman[11] (NM) and the landmark Aalen-Johansen estimator (LMAJ). The methods of Allignol et al.[8] and de Uña Àlvarez & Meira-Machado[9] are not available for reversible illness-death models and were therefore not included in this comparison. Table 3 shows results for $P_{11}(s,t)$ and $P_{21}(s,t)$. The Aalen-Johansen estimator outperforms the non-Markov estimators when the model is actually Markov. Interestingly, in many cases where departure from Markovianity is modest, the Aalen-Johansen estimator does show a moderate bias, but the smaller variance still results in smaller RMSE. Coverage, however, is noticeably below the nominal level of 95%, because of the bias. As in the first simulation study, both NM and LMAJ are unbiased with good coverage, and the latter consistently shows a somewhat lower RMSE than the former.

Reversible four-state extended illness-death model

The third set of simulations assessed the performance of the same three estimators, AJ, NM and LMAJ, in a reversible four-state extended illness-death model with frailty. States 1, 2 and 3 represent progressively serious illness states with transitions back and forth between 1 and 2 and between 2 and

				AJ			NM			LMAJ		
(β_1, β_2)	n	$\ell \to m$	Truth	Bias	RMSE	Cov	Bias	RMSE	Cov	Bias	RMSE	Cov
(1, 1) (Markov)	200	$1 \to 1$	0.3084	0.008	4.472	94.5	0.195	5.291	94.3	0.031	5.066	94.8
		$2 \to 1$	0.1506	0.000	3.075	94.2	0.114	5.477	92.3	-0.041	5.329	93.2
	500	$1 \to 1$	0.3084	0.005	2.822	95.1	0.057	3.288	94.8	0.008	3.171	95.0
		$2 \to 1$	0.1506	-0.021	1.930	94.5	0.034	3.509	93.9	-0.031	3.365	94.7
(1.5, 1)	200	$1 \to 1$	0.2140	1.697	4.265	93.2	0.055	4.463	93.8	-0.039	4.430	94.1
		$2 \to 1$	0.1669	-4.100	5.010	60.9	0.284	6.115	92.5	0.069	5.883	94.2
	500	$1 \to 1$	0.2140	1.720	2.955	90.8	0.037	2.759	94.9	0.020	2.667	95.0
		$2 \to 1$	0.1669	-4.147	4.505	33.6	0.045	3.834	94.1	-0.065	3.723	94.5
(1, 0.5)	200	$1 \to 1$	0.3230	-0.846	4.565	94.2	0.141	5.136	94.2	0.014	4.984	94.8
		$2 \to 1$	0.1299	2.437	4.037	92.3	0.250	6.334	90.2	0.066	6.098	94.4
	500	$1 \to 1$	0.3230	-0.906	2.972	94.1	0.058	3.261	94.5	-0.015	3.167	94.5
		$2 \to 1$	0.1299	2.449	3.189	81.4	0.062	3.990	92.6	-0.024	3.828	93.6
(1.5, 0.5)	200	$1 \to 1$	0.2326	0.513	3.983	94.8	0.107	4.330	94.3	-0.012	4.215	94.7
		$2 \to 1$	0.1452	-2.009	3.504	82.1	0.300	7.193	89.6	0.033	6.883	93.8
	500	$1 \to 1$	0.2326	0.531	2.525	94.9	0.044	2.740	94.6	-0.001	2.676	94.7
		$2 \to 1$	0.1452	-2.041	2.725	73.9	0.052	4.476	93.1	-0.050	4.277	94.2

Table 3. Bias, root mean squared error (RMSE) and coverage (all $\times 100$) of the Aalen-Johansen (AJ), the method of Titman[11] (NM) and the new proposed landmark Aalen-Johansen method (LMAJ) for estimating $P_{\ell m}(5, 15)$, for different values of ℓ and m. Truth refers to the true value of $P_{\ell m}(5, 15)$.

3, and transitions between each of states 1, 2 and 3 and a death state 4. Transition intensities were taken as $\alpha_{jk}Z$, with Z a gamma frailty with unit mean and variance 0 (no frailty, so Markov), 1 and 2, with $\alpha_{12} = \alpha_{23} = 0.20$, $\alpha_{21} = \alpha_{32} = 0.10$, and $\alpha_{14} = 0.06$, $\alpha_{24} = 0.09$, $\alpha_{34} = 0.12$. $M = 5000$ data sets of size $n = 500$ were generated with independent right censoring from a uniform distribution on $(5, 40)$. Table 4 reports bias, root mean squared error (RMSE) and coverage for estimating $P_{11}(s,t)$, $P_{12}(s,t)$, $P_{21}(s,t)$ and $P_{22}(s,t)$ for $s = 5$ and $t = 15$. The overall picture is similar to that of the previous simulation studies. For the Markov model (frailty variance equal to 0) the Aalen-Johansen estimator performs best, with considerably smaller RMSE compared to the other robust estimators. For the non-Markov case (frailty variance 1 and 2), the Aalen-Johansen estimator is biased, leading to an increased RMSE and unacceptable coverage, increasingly so for increasing frailty variance. Both NM and LMAJ perform adequately in terms of bias and coverage, also in the non-Markov case. The LMAJ estimator consistently has a somewhat smaller RMSE compared to NM.

Application

We further compare the landmark Aalen-Johansen method with the Aalen-Johansen method in data from a clinical trial in breast cancer patients, conducted by the European Organization for Research and Treatment of Cancer (EORTC trial 10854). The objective of the trial was to study whether a short intensive course of perioperative chemotherapy yields better overall survival than surgery alone. The trial included 2795 patients with early breast cancer, who underwent either radical mastectomy or breast conserving therapy before being randomized. Patients were randomized to either perioperative chemotherapy or no perioperative chemotherapy. Results of the trial were reported in [19;20]. In this analysis we consider the same 2687 eligible patients that were also studied in earlier analyses [15;21;22]. There it was noted that patients with early local recurrence had a higher transition rate from local recurrence than patients with later local recurrence, pointing to a possible violation of the Markov assumption.

We consider a multi-state model with states “Surgery” (state 1), “Local Recurrence” (state 2), and “Death” (state 3). The multi-state model is an irreversible illness-death model, with transitions from Surgery to Local Recurrence and Death, and from Local Recurrence to Death. Of 2687 patients, 84 patients died directly, without prior local recurrence; 1061 experienced a local recurrence, of which 645 died afterwards. The remaining patients were censored, 1542 in state 1 and 416 in state 2. The total number of deaths observed was 729.

Figure 1 shows estimated curves of the conditional probabilities of being alive with local recurrence at time t, and of having died by time t, conditional on being alive without local recurrence at 2 years, for both randomized treatment groups. In the notation of this paper, these are the transition probabilities $P_{\ell m}(s,t)$, $t \geq s$, for $\ell = 1$, $m = 2$ and 3, and $s = 2$ years. Two different estimates are considered; the

				AJ			NM			LMAJ		
Variance	n	$\ell \to m$	Truth	Bias	RMSE	Cov	Bias	RMSE	Cov	Bias	RMSE	Cov
0 (Markov)	500	$1 \to 1$	0.1893	-0.004	2.571	94.7	0.016	3.317	94.3	-0.023	3.241	94.8
		$1 \to 2$	0.1680	0.018	2.440	94.3	0.013	3.275	94.0	-0.014	3.231	94.4
		$2 \to 1$	0.1338	-0.008	1.989	94.5	0.116	3.679	93.2	0.020	3.545	94.9
		$2 \to 2$	0.1455	-0.006	2.148	94.9	0.054	3.679	94.2	0.017	3.626	95.2
1	500	$1 \to 1$	0.4853	-4.154	5.228	72.7	0.118	3.650	94.8	0.060	3.483	95.1
		$1 \to 2$	0.1756	0.921	2.513	94.5	0.002	2.821	94.7	-0.026	2.726	94.8
		$2 \to 1$	0.1397	9.848	10.160	1.6	0.018	4.209	92.8	-0.042	4.050	94.5
		$2 \to 2$	0.2305	-4.279	4.955	57.7	0.127	5.099	94.5	0.004	4.973	95.0
2	500	$1 \to 1$	0.6358	-4.244	5.208	69.3	0.122	3.292	94.3	-0.127	3.156	94.6
		$1 \to 2$	0.1363	1.365	2.477	92.3	0.065	2.409	94.0	0.145	2.309	94.8
		$2 \to 1$	0.1354	15.000	15.327	0.1	-0.074	4.777	93.2	-0.066	4.595	93.8
		$2 \to 2$	0.2605	-7.110	7.643	28.3	0.257	6.035	94.2	0.047	5.841	95.5

Table 4. Bias, root mean squared error (RMSE) and coverage (all $\times 100$) of the Aalen-Johansen (AJ), the method of Titman[11] (NM) and the new proposed landmark Aalen-Johansen method (LMAJ) for estimating $P_{\ell m}(5, 15)$, for different values of ℓ and m. Truth refers to the true value of $P_{\ell m}(5, 15)$.

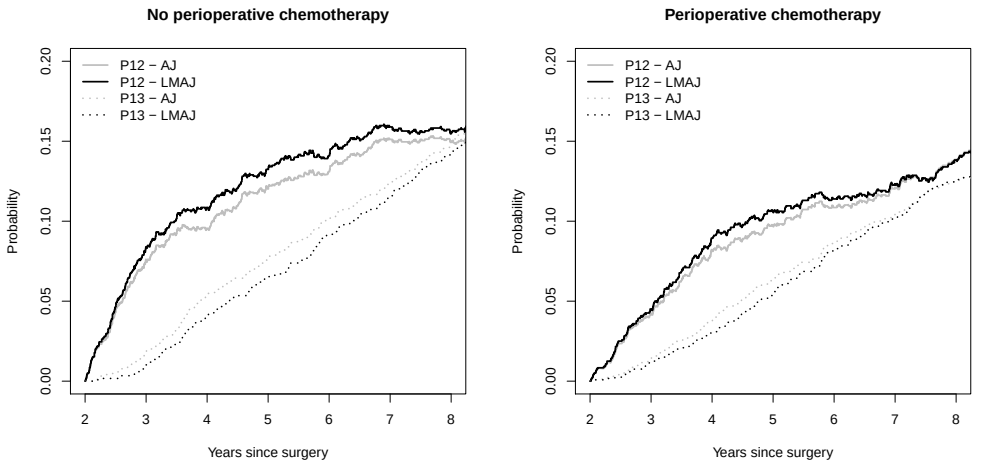

Figure 1. Estimated transition probabilities $P(X(t) = 2 \mid X(2) = 1)$ and $P(X(t) = 3 \mid X(2) = 1)$ for the two randomized treatment groups, with estimates based on (a) Aalen-Johansen, (b) the landmark Aalen-Johansen estimator.

Aalen-Johansen (AJ) estimator that is valid only when the Markov assumption is satisfied, and the robust landmark Aalen-Johansen (LMAJ) estimator. It is seen from Figure 1 that, conditional on being alive without local recurrence at 2 years, both the probabilities of being alive with local recurrence and having died are somewhat higher for the no perioperative chemotherapy group. Comparing the two different estimators, it is seen that for both treatment groups the robust estimator results in a somewhat more optimistic picture; the estimated probability of having died is lower for LMAJ compared to AJ, and the estimated probability of being alive with local recurrence is higher for LMAJ compared to AJ. The estimated probability of being alive without local recurrence (not shown) is very similar for LMAJ and AJ.

Discussion

In this paper we showed that a simple and intuitive procedure combining landmarking (subsampling) and the Aalen-Johansen estimator of the (conditional) state occupation probabilities yields a consistent estimator of transition probabilities in general non-Markov multi-state models. The method is comparable to the estimator of de Uña Àlvarez & Meira-Machado[9] (conditional Pepe) with respect to bias and RMSE, but can be used beyond the irreversible illness-death model. It shows a slight improvement in terms of RMSE to the method of Titman[11]. In addition, unlike the non-Markov estimator of Titman[11],

the transition probability estimators $\widehat{P}_{\ell m}^{\text{LMAJ}}(s,t)$, when summed over all states m, add up to one. Interestingly, our simulation studies indicated that on occasion the Aalen-Johansen estimator outperforms the non-Markov estimators when the multi-state model exhibits modest deviations from Markovianity. In such cases the modest bias does not weigh against the smaller variance of the Aalen-Johansen estimator. Coverage of the Aalen-Johansen estimator is too low, however. If departure from Markovianity increases, the robust estimators clearly are to be preferred.

Titman[11] proposes the use of pseudo-observations when interest is in the effect of covariates on the transition probabilities. This indeed provides a useful alternative to fitting regression models to the transition intensities in a multi-state model. The LMAJ estimator can also be used for this purpose. In fact, this approach has been used to model the expected length of stay in a given state in a multi-state models, based on the LMAJ estimator, also in non-Markovian models[23].

The landmark Aalen-Johansen method has been implemented (function `LMAJ`) in the latest version (0.2.9) of the **mstate** package[24] in R.

Acknowledgement

The authors are grateful to Andrew Titman for providing code implementing his non-Markov estimator. The European Organisation for Research and Treatment of Cancer (EORTC) is gratefully acknowledged for making available the data of EORTC trial 10854.

References

1. Andersen PK and Keiding N. Multi-state models for event history analysis. *Stat Methods Med Res* 2002; 11: 91–115.
2. Putter H, Fiocco M and Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* 2007; 26: 2389–2430.
3. Meira-Machado L, de Uña Álvarez J, Cadarso-Suárez C et al. Multi-state models for the analysis of time-to-event data. *Stat Methods Med Res* 2009; 18: 195–222.
4. Beyersmann J, Schumacher M and Allignol A. *Competing Risks and Multistate Models with R*. Springer, New York, 2012.
5. Geskus RB. *Data Analysis with Competing Risks and Intermediate States*. Boca Raton: Chapman & Hall / CRC, 2016.
6. Aalen OO and Johansen S. An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand Stat Theory Appl* 1978; 5: 141–150.

7. Meira-Machado L, de Uña Álvarez J and Cadarso-Suárez C. Nonparametric estimation of transition probabilities in a non-Markov illnessdeath model. *Lifetime Data Anal* 2006; 12: 325–344.
8. Allignol A, Beyersmann J, Gerds T et al. A competing risks approach for nonparametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Anal* 2014; 20: 495–513.
9. de Uña Álvarez J and Meira-Machado L. Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study. *Biometrics* 2015; 71: 364–375.
10. Pepe MS. Inference for events with dependent risks in multiple endpoint studies. *J Am Stat Assoc* 1991; 86: 770–778.
11. Titman AC. Transition probability estimates for non-Markov multi-state models. *Biometrics* 2015; 71: 1034–1041.
12. Datta S and Satten GA. Validity of the Aalen–Johansen estimators of stage occupation probabilities and Nelson–Aalen estimators of integrated transition hazards for non-Markov models. *Stat Probab Lett* 2001; 55: 403–411.
13. Glidden DV. Robust inference for event probabilities with non-Markov event data. *Biometrics* 2002; 58: 361–368.
14. Anderson JR, Cain KC and Gelber RD. Analysis of survival by tumor response. *J Clin Oncol* 1983; 1: 710–719.
15. van Houwelingen H and Putter H. *Dynamic Prediction in Clinical Survival Analysis*. Chapman & Hall / CRC, Boca Raton, 2012.
16. Putter H. *Handbook of Survival Analysis*, chapter 21. Landmarking. Chapman & Hall/CRC, Boca Raton, 2013. pp. 441–456.
17. Spitoni C, Verduijn M and Putter H. Estimation and asymptotic theory for transition probabilities in Markov renewal multi-state models. *Int J Biostat* 2012; 8: 23.
18. Pepe MS and Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. *J Am Stat Assoc* 1993; 88: 811–820.
19. Clahsen PC, van de Velde CJH, Julien JP et al. Improved local control and disease-free survival after perioperative chemotherapy for early-stage breast cancer. *Journal of Clinical Oncology* 1996; 14: 745–753.
20. van der Hage JA, van de Velde CJH, Julien JP et al. Improved survival after one course of perioperative chemotherapy in early breast cancer patients: long-term results from the European Organization for Research and Treatment of Cancer (EORTC) trial 10854. *European Journal of Cancer* 2001; 37: 2184–2193.
21. Putter H, van der Hage J, de Bock GH et al. Estimation and prediction in a multi-state model for breast cancer. *Biometrical Journal* 2006; 48: 366–380.
22. Putter H and van Houwelingen JC. Frailties in multi-state models: Are they identifiable? Do we need them? *Stat Meth Med Res* 2015; 24: 675–692.
23. Klinten Grand M and Putter H. Regression models for expected length of stay. *Stat Med* 2016; 35: 1178–1192.
24. de Wreede LC, Fiocco M and Putter H. The mstate package for estimation and prediction in non-and semi-parametric multi-state and competing risks models. *Comput Methods Programs Biomed* 2010; 99: 261–274.

Appendix: Consistency of the landmark Aalen-Johansen estimator

Here we show that if $P(\tilde{X}(s) = \ell) > 0$ and the same conditions as needed for consistency of the Aalen-Johansen estimator of the state occupation probabilities[12] are satisfied, the landmark Aalen-Johansen estimator will also be consistent.

Fix ℓ and s. From the original multi-state process $\tilde{X}(t)$ with state space $\{1, \ldots, K\}$ define the coupled multi-state process $X^\star(t)$ with enlarged state space $\{-K, \ldots, -1, 1, \ldots, K\}$ by $X^\star(t) = \tilde{X}(t)$ for $t < s$ and $X^\star(t) = (2 \cdot \mathbf{1}\{\tilde{X}(s) = \ell\} - 1) \cdot \tilde{X}(t)$, for $t \geq s$. In words, $X^\star(t)$ follows the original multi-state model $\tilde{X}(t)$ until just before time s, while for $t \geq s$, $X^\star(t)$ follows either $\tilde{X}(t)$, if $\tilde{X}(s) = \ell$, or diverges to $-\tilde{X}(t)$, if $\tilde{X}(s) \neq \ell$. Note that the process $X^\star(\cdot)$ is not Markov even in case $X(\cdot)$ is Markov: for any $t > s$, $X^\star(t)$ depends on the past through $X^\star(s)$. Since state $m \geq 1$ at time $t > s$ can be reached only if $\tilde{X}(s) = \ell$, this artificial multi-state model has, for $t > s$, $m \geq 1$:

$$P^\star_m(t) = P(X^\star(t) = m) = P(\tilde{X}(t) = m, \tilde{X}(s) = \ell),$$

so that the transition probability of interest can be written as

$$P_{\ell m}(s,t) = P(\tilde{X}(t) = m \,|\, \tilde{X}(s) = \ell) = \frac{P^\star_m(t)}{P^\star_\ell(s)},$$

a ratio of two state occupation probabilities of the coupled multi-state process. By the results in Datta & Satten[12], the Aalen-Johansen estimators $\widehat{P}^\star_m(t)$ and $\widehat{P}^\star_\ell(s)$ of the state occupation probabilities $P^\star_m(t)$ and $P^\star_\ell(s)$ are consistent, and hence their ratio consistently estimates the transition probability of interest if $P^\star_\ell(s) > 0$.

The ratio of the Aalen-Johansen estimates of state occupation probabilities can be written as

$$\begin{aligned}
\frac{\widehat{P}^\star_m(t)}{\widehat{P}^\star_\ell(s)} &= \frac{\left[\hat{\pi}^\star(0) \prod_{0<u\leq t} \left(\mathbf{I} + d\widehat{\mathbf{\Lambda}}^\star(u)\right)\right]_m}{\left[\hat{\pi}^\star(0) \prod_{0<u\leq s} \left(\mathbf{I} + d\widehat{\mathbf{\Lambda}}^\star(u)\right)\right]_\ell} \\
&= \frac{\sum_j \left[\hat{\pi}^\star(0) \prod_{0<u\leq s} \left(\mathbf{I} + \Delta\widehat{\mathbf{\Lambda}}^\star(u)\right)\right]_j \left[\prod_{s<u\leq t} \left(\mathbf{I} + \Delta\widehat{\mathbf{\Lambda}}^\star(u)\right)\right]_{jm}}{\left[\hat{\pi}^\star(0) \prod_{0<u\leq s} \left(\mathbf{I} + \Delta\widehat{\mathbf{\Lambda}}^\star(u)\right)\right]_\ell} \\
&= \frac{\left[\hat{\pi}^\star(0) \prod_{0<u\leq s} \left(\mathbf{I} + \Delta\widehat{\mathbf{\Lambda}}^\star(u)\right)\right]_\ell \left[\prod_{s<u\leq t} \left(\mathbf{I} + \Delta\widehat{\mathbf{\Lambda}}^\star(u)\right)\right]_{\ell m}}{\left[\hat{\pi}^\star(0) \prod_{0<u\leq s} \left(\mathbf{I} + \Delta\widehat{\mathbf{\Lambda}}^\star(u)\right)\right]_\ell} \\
&= \left[\prod_{s<u\leq t} \left(\mathbf{I} + \Delta\widehat{\mathbf{\Lambda}}^\star(u)\right)\right]_{\ell m},
\end{aligned}$$

where $\hat{\pi}^\star(0)$ and $\widehat{\mathbf{\Lambda}}^\star$ are, respectively, the initial state proportions and the estimated cumulative hazards matrix of $X^\star$, and all matrix products are over the observed transition times u. The third equality follows because for all $j \neq \ell$, $\left[\prod_{s<u\leq t}\left(\mathbf{I}+\Delta\widehat{\mathbf{\Lambda}}^\star(u)\right)\right]_{jm}$ is zero, because just before the first event time after s, everyone is either at state ℓ, or has been redirected to a negative state, from which state $m \geq 1$ cannot be reached.

The last step we need for proving the theorem is to show that

$$\left[\prod_{s<u\leq t}\left(\mathbf{I}+\Delta\widehat{\mathbf{\Lambda}}^\star(u)\right)\right]_{\ell m} = \hat{\pi}^{(\mathrm{LM})}(s)\prod_{s<u\leq t}\left(\mathbf{I}+\Delta\widehat{\mathbf{\Lambda}}^{(\mathrm{LM})}(u)\right).$$

This follows by noting that the matrices $\mathbf{I}+\Delta\widehat{\mathbf{\Lambda}}^\star(u)$ on the left hand side are $2K \times 2K$ diagonal block matrices, consisting of two blocks representing the positive states and the negative ones, with no interaction between them for $u > s$. The only relevant $K \times K$ sub-matrices are those representing the positive states. For these sub-matrices, note that the counting and at-risk processes $N^\star_{ijk}(t)$ and $Y^\star_{ij}(t)$, defining the $\widehat{\mathbf{\Lambda}}^\star(t)$ used in the Aalen-Johansen estimator, are given by

$$N^\star_{ijk}(t) = N_{ijk}(t)1\{X_i(s)=\ell\}, \text{ and } Y^\star_{ij}(t) = Y_{ij}(t)1\{X_i(s)=\ell\},$$

because if $X_i(s) \neq \ell$, then $X^\star_i(t)$ would be negative for $t \geq s$. So for $t > s$, we have that $\overline{N}^{(\mathrm{LM})}_{jk}(t) = \overline{N}^\star_{jk}(t)$, $\overline{Y}^{(\mathrm{LM})}_{j}(t) = \overline{Y}^\star_{j}(t)$, and hence $\Delta\widehat{\mathbf{\Lambda}}^{(\mathrm{LM})}(t) = \left[\Delta\widehat{\mathbf{\Lambda}}^\star(t)\right]_{1,\ldots,K;1,\ldots,K}$. This concludes the proof.