Econometrics and Statistics 25 (2023) 110–124

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Statistical inference for state occupation and transition probabilities in non-Markov multi-state models subject to both random left-truncation and right-censoring

Alexandra Nießl[a,d,*], Arthur Allignol[b], Jan Beyersmann[a], Carina Mueller[c]

[a] Institute of Statistics, Ulm University, Helmholtzstraße 20, 89069 Ulm, Germany
[b] Daiichi Sankyo Europe GmbH, Luitpoldstraße 1, 85276 Pfaffenhofen an der Ilm, Germany
[c] Metronomia Clinical Research GmbH, Paul-Gerhardt-Allee 42, 81245 München, Germany
[d] Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Str. 65, 88397 Biberach, Germany

ARTICLE INFO

Article history:
Received 13 October 2020
Revised 28 September 2021
Accepted 30 September 2021
Available online 21 October 2021

Keywords:
Nelson-Aalen estimator
Wild bootstrap
Hospital epidemiology
Partly conditional transition rate
Methicillin-resistant *staphylococcus aureus*

ABSTRACT

The Aalen-Johansen estimator generalizes the Kaplan-Meier estimator for independently left-truncated and right-censored survival data to estimate the transition probability matrix of a time-inhomogeneous Markov model with finite state space. Such multi-state models have a wide range of applications for modelling complex courses of a disease over the course of time, but the Markov assumption may often be in doubt. If censoring is entirely unrelated to the multi-state data, it has been suggested that the Aalen-Johansen estimator, standardized by the initial empirical distribution of the multi-state model, still consistently estimates the state occupation probabilities. Recently, this approach has been extended to transition probabilities using landmarking, which is, inter alia, useful for dynamic prediction. However, there have been recent concerns about the mathematical arguments leading to the former result. These findings are complemented in three ways. Firstly, a rigorous proof of consistency of the Aalen-Johansen estimator for state occupation probabilities, on which also correctness of the landmarking approach hinges, is presented correcting and simplifying the earlier result. Secondly, delayed study entry is a common phenomenon in observational studies, and the earlier results are extended to multi-state data also subject to left-truncation. Thirdly, the rigorous proof is suggestive of wild bootstrap resampling. Studying wild bootstrap is motivated by the fact that it is desirable to have a technique that works for models where left-truncation and right-censoring need not be entirely random, then requiring a Markov assumption, and that may still perform reasonably with non-Markov models subject to random left-truncation and right-censoring. The developments for left-truncation are motivated by a prospective observational study on the occurrence and the impact of a multi-resistant infectious organism in patients undergoing surgery. Both the real data example and simulation studies are presented. The case for wild bootstrapping is illustrated for event-driven trials, where the data are censored once a prespecified number of events have been observed.

* Corresponding author.
E-mail address: alexandra.niessl@uni-ulm.de (A. Nießl).

https://doi.org/10.1016/j.ecosta.2021.09.008

1. Introduction

Aalen and Johansen (1978) developed an estimator of the transition probability matrix of a non-homogeneous Markov multi-state model for independently left-truncated and right-censored data. These models are useful for studying complex courses of a disease over the course of time and applications in medical research include oncology (Schmoor et al., 2013), cardiology (Gasperoni et al., 2017), gastroenterology (Jepsen et al., 2015), orthopaedics (Gillam et al., 2012) or hospital epidemiology (Munoz-Price et al., 2016). However, the Markov assumption may regularly be in doubt in applications. In our motivating data example, De Angelis et al. (2011) investigated the occurrence and the impact of Methicillin-resistant *staphylococcus aureus* (MRSA) infection in hospital compared to patients only colonized with MRSA, using an illness-death multi-state model. Violations of the Markov assumption arise if the time of MRSA infection affects the hazard of end of hospital stay. See also Andersen and Keiding (2002) for a clear practical discussion of a non-Markov multi-state model.

The Markov assumption enters the technical developments in Aalen and Johansen (1978) (see also Andersen et al., 1993) in that it implies a particularly handy form of the intensities of the counting processes of observed transitions between any two states of the model. Save for the at-risk processes, these intensities are non-random and equal to the usual transition hazards. This is in contrast to the non-Markov case where the intensities will also be random through dependence on the past. For instance, in the MRSA data, such dependence is present if the time of infection affects the end-of-stay hazard of an infected patient.

For non-Markov models and complete observations, Andersen et al. (1993, Section IV.4.1.4) showed that the Aalen-Johansen estimator, standardized by the multinomial estimator of the initial distribution of the multi-state model, results in the usual multinomial estimator of the unconditional state occupation probabilities. Later, Datta and Satten (2001) suggested that this approach still yields a consistent estimator of the state occupation probabilities based on right-censored observations provided that censoring is entirely random. However, recent authors (Müller, 2015; Overgaard, 2019; Maltzahn et al., 2020) have questioned the validity of the mathematical arguments in Datta and Satten (2001), which potentially also affects the recent extension of Putter and Spitoni (2016) to a landmark Aalen-Johansen estimator of the transition probabilities in randomly right-censored non-Markov multi-state models. Their estimator is based on Aalen-Johansen estimates of the state occupation probabilities computed on subsamples of the data. Consistency of the landmark Aalen-Johansen estimator then follows provided that the Aalen-Johansen estimator of the state occupation probabilities is consistent.

The worry is that the arguments of Datta and Satten (2001) are compromised by their use of non-random intensities, which do *not* apply in non-Markov models. The issue is this: Datta and Satten built on the result of Andersen et al. (1993) for complete data. Their idea was to show that the multivariate Nelson-Aalen estimator, of which the Aalen-Johansen estimator is a transformation, consistently estimates the same limit both in the completely observed and in the randomly right-censored case. Consistency of estimating the state occupation probabilities via the Aalen-Johansen estimator then follows from the continuous mapping theorem, although this approach was not taken by Datta and Satten (2001) who used an inverse probability of censoring argument. For the multivariate Nelson-Aalen estimator, Datta and Satten (2001) started with complete data, used martingale methods similar to Aalen and Johansen (1978) for the Markov case and then transferred results to the right-censored case using inverse probability of censoring weights (Horvitz and Thompson, 1952). However, their arguments used intensities that were, save for the at-risk processes, non-random (see their Equation (A.5)), and use of inverse probability of censoring weights makes the arguments unnecessarily complicated.

Therefore, Overgaard (2019) avoided the use of martingale arguments and showed consistency of the Aalen-Johansen estimator for state occupation probabilities using the perhaps less familiar concept of interval functions (Gill and Johansen, 1990; Dudley and Norvaiša, 2011). In contrast, we will use martingale methods, working, however, with the proper intensities. Being at the methodological core of survival and event history analysis (Aalen et al., 2008), martingale methodology both allows us to incorporate left-truncation and it lends itself to wild bootstrap resampling which we may find preferable to simple drawing with replacement from the data. At this stage, accounting for left-truncation is achieved via the usual modification of the risk sets that may also increase as a consequence of delayed entry. Left-truncation will also impact the data used in a landmark dataset as explained later.

Incorporating left-truncation is motivated by a prospective observational study in the field of hospital epidemiology on the occurrence and the impact of a multi-resistant infectious organism in patients undergoing surgery. The usual right-censoring is not much of an issue in hospital epidemiology, but delayed study entry, i.e., left-truncation may very well be (Beyersmann et al., 2011). Our motivating study considers a prospective cohort of patients colonized with MRSA. The time scale of interest is time since hospital admission, and patients may have a delayed study entry if a positive laboratory result is only available some time after admission. Left-truncation is a common phenomenon in observational studies (Bluhmki et al., 2017) and an extension of the findings mentioned above would be generally useful.

Wild bootstrapping is motivated by the desire to obtain a resampling procedure where, in the words of Aalen and Johansen (1978), the censoring process may, “in almost arbitrary ways”, depend on the past, then requiring a Markov multi-state model, and that we may expect to still perform reasonably with non-Markov data subject to random left-truncation and random right-censoring. We will illustrate the former for non-random censoring triggered by having observed a prespecified number of events. This type of censoring is common in pharmaceutical trials and it leads to dependent time-to-event trajectories.

The basic idea of our novel proof is that all individual transition-specific counting processes have a compensator regardless of the Markov assumption. However, in the non-Markov case, the individual intensities truly depend on the past individual trajectories through the multi-state model and, hence, are random quantities not only through the at-risk processes (see Equation (6) below). However, in an i.i.d. setting with entirely random left-truncation and right-censoring that is unrelated to the underlying multi-state model, we will find that the transition-specific Nelson-Aalen estimator averages over these random intensities, obtaining the same limit as in the complete data case. The assumption that restrictions on the observable data are entirely random, i.e., unrelated to the process of interest, is well known from the recurrent events literature (e.g. Andersen et al., 2019; Cook and Lawless, 2007, e.g. Section 7.2.2; Aalen et al., 2008, Section 8.1) and can be seen as the price to pay to handle non-Markov models.

The remainder of the paper is organized as follows: The methodological developments are in Section 2, simulations and the analysis of the MRSA data are reported in Section 3 and a discussion is in Section 4. Proofs are in the Appendix.

2. Methods

We begin by introducing notation and the Nelson-Aalen and Aalen-Johansen estimators in possibly non-Markov multi-state models. Estimating state occupation probabilities will be considered in Subsection 2.1, estimating transition probabilities in Subsection 2.2, and Subsection 2.3 will consider resampling with an emphasis on the wild bootstrap relying on a martingale representation from Subsection 2.1.

Let $(X(t))_{t\geq 0}$ be a stochastic process with state space $\{0, 1, \dots, K\}$ with right-continuous sample paths having left-hand limits. This multi-state process may be non-Markov with a non-degenerate initial distribution. The first aim is to estimate the state occupation probabilities

$$P(X(t) = m),\ m \in \{0, 1, \dots, K\}, t \in [0, \tau] \tag{1}$$

allowing for both random left-truncation and right-censoring. In a second step, we will also estimate transition probabilities

$$P_{lm}(s, t) = P(X(t) = m \mid X(s) = l),\ l, m \in \{0, 1, \dots, K\}, s \leq t \in [0, \tau], \tag{2}$$

using an estimator of $P(X(t) = m)$ in a landmark (sub-) data set that accounts for conditioning on $X(s) = l$. Landmarking for such a purpose has been proposed by Allignol et al. (2014) for the special case of an illness-death model and later, for general multi-state models, by Putter and Spitoni (2016). To this end, define the partly conditional $l \to m$ transition rate (Pepe and Cai, 1993)

$$\alpha_{lm}(t) = \lim_{\Delta t \searrow 0} P(X(t + \Delta t) = m \mid X(t-) = l),\ l, m \in \{0, 1, \dots, K\}, l \neq m, \tag{3}$$

which we assume to exist, with cumulative quantities $A_{lm}(t) = \int_0^t \alpha_{lm}(u)\, \mathrm{d}u$.

We assume that observation of X is subject to random left-truncation by L and right-censoring by C. Denote the event of study entry, i.e., X reaches an absorbing state after L, by Z. Given study entry, consider i.i.d. data $(X_i(t))_{t\in(L_i, C_i\wedge T_i]}$, $i = 1, \dots, n$, of n individuals under study, where T_i is i's time until absorption and $\wedge$ denotes the minimum. The random censoring assumption enters this set-up in that, e.g., event-driven (non-random) censoring investigated later leads to statistical dependence across units. Let $\mathcal{G}(t)$ denote the self-exciting filtration of the observed data $(X_i(t))_{t\in(L_i, C_i\wedge T_i]}$, $i = 1, \dots, n$.

Define the individual counting process

$$N_{i;lm}(t) = \#\text{ of observed } l \to m \text{ transitions of } i \text{ in } [0, t], \tag{4}$$

$l \neq m$, and the individual at-risk process

$$Y_{i;l}(t) = \mathbf{1}(X_i(t-) = l, L_i < t \leq C_i), \tag{5}$$

such that

$$N_{i;lm}(t) = \sum_{u\in[0,t]} Y_{i;l}(u) \cdot \mathbf{1}(X_i(u) = m),$$

where the sum in the previous display is w.r.t. all unique, observed transition times u. The counting process of observed $l \to m$ transitions is $N_{lm}(t) = \sum_{i=1}^n N_{i;lm}(t)$ and the at-risk process for state l is $Y_l(t) = \sum_{i=1}^n Y_{i;l}(t)$. Also introduce $J_l(t) = \mathbf{1}(Y_l(t) > 0)$. We assume that the $N_{i;lm}$'s have absolutely continuous compensators with respect to $\mathcal{G}$, such that

$$M_{lm}(t) = N_{lm}(t) - \int_0^t \sum_{i=1}^n Y_{i;l}(u) \cdot \alpha_{i;lm}(u|\mathcal{G}(u-))\, \mathrm{d}u, \tag{6}$$

is a mean zero martingale with respect to $\mathcal{G}$. Here, as usual, we assume that both the counting processes and the integrals on the right hand side of (6) are finite. If $(X(t))_{t\geq 0}$ is time-inhomogeneous Markov, the intensity $\alpha_{i;lm}(t|\mathcal{G}(t-))$ will be equal to $\alpha_{lm}(t)$ from (3), but in general the intensity will be a random quantity through its dependence on the past.

The Nelson-Aalen estimator is

$$\hat{A}_{lm}(t) = \int_0^t \frac{J_l(u)}{Y_l(u)}\, \mathrm{d}N_{lm}(u), \tag{7}$$

and the Aalen-Johansen estimator is, using product integral notation,

$$\prod_{u\in(0,t]} \left(\mathbf{I} + \mathrm{d}\hat{\mathbf{A}}(u)\right), \tag{8}$$

where $\mathbf{I}$ is the $(K+1)\times(K+1)$ identity matrix. The matrix $\hat{\mathbf{A}}(t)$ has non-diagonal entries $\hat{A}_{lm}(t)$ and diagonal entries are such that the sum of each row equals zero.

So far, we have not established what quantities (7) and (8) estimate. Subsection 2.1 moves from (7) to (8) for estimating state occupation probabilities, while Subsection 2.2 uses a landmark version of (8) for estimating transition probabilities. Throughout, we will allow for both random left-truncation and right-censoring. Left-truncation will imply some subtleties as outlined below. Loosely speaking, the assumption that left-truncation and right-censoring are entirely unrelated to the multi-state process ensures that the average of the intensities $\alpha_{i;lm}(t|\mathcal{G}(t-))$ approaches the partly conditional transition rate. Under this assumption, the Nelson-Aalen estimator presents itself as a natural estimator of the cumulative partly conditional transition rate, because

$$\begin{aligned}\alpha_{lm}(t)\,\mathrm{d}t &= P(X(t+\,\mathrm{d}t) = m\,|\,X(t-) = l)\\ &= P(X(t+\,\mathrm{d}t) = m\,|\,X(t-) = l,\ L < t \le C).\end{aligned}$$

The last line conditions on being in state l *and* being under observation, while the length $\mathrm{d}t$ of the infinitesimal time interval is so short that an $l \to m$-transition in this time interval will be observed. Hence, under the assumption of random left-truncation and right-censoring, it is natural to estimate $\alpha_{lm}(t)\,\mathrm{d}t$ by the number of observed $l \to m$-transitions at time t divided by the number of units under observation in state l just prior time t, cf. (7). Section 2.1 provides a rigorous statement.

2.1. The Aalen-Johansen estimator of the state occupation probabilities

The following result is similar to the classical strong consistency theorem of the multivariate Nelson-Aalen estimator for time-inhomogeneous Markov multi-state models (Andersen et al., 1993, Theorem IV.1.1).

Theorem 2.1. *Let $t \in [0,\tau]$ and $l, m \in \{0,1,\dots,K\}, l \neq m$. Assume there exists a function k, $\int_0^t k(u)\,\mathrm{d}u < \infty$, such that*

$$\alpha_{i;lm}(u|\mathcal{G}(u-)) \le k(u) \text{ on}[0,t]\text{with probability } 1, \tag{9}$$

for all $i = 1,\dots n$. Furthermore, as $n\to\infty$, assume that

$$\int_0^t \frac{J_l(u)}{Y_l(u)} k(u)\,\mathrm{d}u \to 0 \text{ in probability}, \tag{10}$$

and

$$\int_0^t (1-J_l(u))k(u)\,\mathrm{d}u \to 0 \text{ in probability}. \tag{11}$$

Then

$$\sup_{u\in[0,t]} \left|\hat{A}_{lm}(u) - A_{lm}(u)\right| \to 0 \text{ in probability}. \tag{12}$$

The proof of Theorem 2.1 in the Appendix uses the proof of Andersen et al. (1993, Theorem IV.1.1) as a template, but with the additional difficulty that $\alpha_{i;lm}(t|\mathcal{G}(t-))$, $i = 1,\dots,n$, are random quantities, unequal to $\alpha_{lm}(t)$. However, assuming i.i.d. multi-state trajectories, these random quantities are i.i.d. too, and their average approaches $\alpha_{lm}(t)$. Before we turn to estimating state occupation probabilities, some remarks on Theorem 2.1 are in place:

1. In the time-inhomogeneous Markov case, the function k can be chosen as
$$k(t) := \sum_{l,m,l\neq m} \alpha_{lm}(t),$$
because the transition hazards are assumed to have finite integrals (Andersen et al., 1993, p. 287).
2. In the presence of left-truncation, the convergence in probability statements are w.r.t. the conditional probability measure given study entry Z from which we sample, see, e.g., Example IV.1.7 of Andersen et al. (1993) and our proofs in the Appendix. Also note that in the absence of left-truncation the main assumption both in our Theorem 2.1 and in the work by Datta and Satten (2001) and as compared to the classical result (Andersen et al., 1993, Theorem IV.1.1) on the Nelson-Aalen estimator is that right-censoring is entirely unrelated to the multi-state process. We will investigate the latter in simulations on wild bootstrap resampling in a non-Markov model where censoring is entirely unrelated and in a time-inhomogeneous Markov model where censoring may depend on the past of the whole cohort.
3. Analogously to the Markov case, a simple condition that implies assumptions (10) and (11) is that the infimum on $[0,\tau]$ of all risk sets Y_l converges in probability to infinity. We refer to Andersen et al. (1993) for an in-depth discussion. This assumption may require reconsidering time 0 in left-truncated studies. E.g., in hospital epidemiology, a natural time

origin is hospital admission. Studies with patients who are colonized with a certain infectious organism upon admission will typically include a substantial proportion of patients with colonization status known at time 0. Other patients will have left-truncated study entries upon arrival of laboratory results (see, e.g., our motivating study De Angelis et al., 2011). In this setting, we may assume the condition to be fulfilled. However, in studies on pregnancy outcomes the natural time origin is conception, but women do not enter observational cohorts before pregnancy detection (e.g., Slama et al., 2014). Time 'zero' in the present context should then be chosen as the earliest time point of detecting pregnancies, around six weeks after the beginning of the menstrual cycle, or perhaps even slightly later, say 7 weeks.
4. In general, left-truncated data may contain information on the multi-state trajectory before study entry, but this information is not used here.

Consistent estimation of the state occupation probabilities now follows from Theorem 2.1.

Theorem 2.2. *Suppose $\hat{p}(0) = (\hat{p}_0(0), \dots, \hat{p}_K(0))$ is a consistent estimator of the initial distribution of the multi-state model,*

$$\hat{p}(0) \to (P(X(0) = 0), \dots, P(X(0) = K)) \text{ in probability as} n \to \infty, \tag{13}$$

and define the $1 \times K$ row vector

$$\hat{p}(t) = \hat{p}(0) \cdot \prod_{u \in (0,t]} \big(\mathbf{I} + \, \mathrm{d}\hat{\mathbf{A}}(u)\big). \tag{14}$$

Then, under the assumptions of Theorem 2.1, we have that

$$\sup_{u \in [0,t]} \left| \hat{p}(u) - (P(X(u) = 0), \dots, P(X(u) = K)) \right| \to 0 \text{ in probability.} \tag{15}$$

We prove Theorem 2.2 in the Appendix. The key assumption is (13). The choice of $\hat{p}(0)$ is trivial, if there is one common initial state, say state 0, with $P(X(0) = 0) = 1$. In the absence of left-truncation, the obvious choice is the multinomial estimate $\hat{p}_j(0) = Y_j(0+)/n$. With left-truncation, it is natural to use $\hat{p}_j(0) = Y_j(0+)/(\sum_l Y_l(0+))$ which estimates $P(X(0) = j, L \le 0)/P(L \le 0) = P(X(0) = j)$, using that left-truncation is random and also requiring $P(L \le 0) > 0$, see our discussion following Theorem 2.1. In this context, we note that the landmark estimator that we discuss next, uses a landmark data set that is constructed such that there is one common state occupied by all individuals at the landmark time.

2.2. The landmark Aalen-Johansen estimator of the transition probabilities

The landmark Aalen-Johansen estimator of Putter and Spitoni (2016) is based on subsampling as are the estimators of de Uña-Álvarez and Meira-Machado (2015) and Titman (2015). The idea is to select individuals that are under observation in a given state at a given time and estimate the state occupation probabilities within this subset. Predating these contributions is the work by Allignol et al. (2014) who derived landmarking for this purpose in the special illness-death model without recovery, already allowing, however, for delayed study entry.

To be precise, let, for any $l \to m$-transition,

$$N_{lm}^{(\mathrm{LM})}(t) = \sum_{i=1}^{n} N_{i;lm}(t) \mathbf{1}(L_i < s < C_i,\, X_i(s) = k), \quad s \le t,$$

be the counting process of the subsample that selects individuals that are observed in state k at time s. $N_{i;lm}(t)$ is defined as in (4). Here, the notation suppresses that only $l \to m$-transitions of individuals i with $X_i(s) = k$ for time s and state k fixed are being counted. Also note that $N_{i;lm}(t)$ counts all of i's $l \to m$-transitions on $[0, t]$. We could consider $N_{i;lm}(t) - N_{i;lm}(s)$ in order to only count transitions after s, but this does not impact the landmark estimator which will only consider the increments of $N_{lm}^{(\mathrm{LM})}$ after s. Similarly, define

$$Y_l^{(\mathrm{LM})}(t) = \sum_{i=1}^{n} Y_{i;l}(t) \mathbf{1}(L_i < s < C_i,\, X_i(s) = k),$$

be the subsample-based at-risk process, cf. (5). The landmark Nelson-Aalen estimator then is

$$\hat{A}_{lm}^{(\mathrm{LM})}(t) = \int_0^t \frac{\mathbf{1}(Y_l^{(\mathrm{LM})}(u) > 0)}{Y_l^{(\mathrm{LM})}(u)} \, \mathrm{d}N_{lm}^{(\mathrm{LM})}(u).$$

Finally, the landmark Aalen-Johansen estimator of

$$P_k(s,t) := (P_{k0}(s,t), P_{k1}(s,t), \dots, P_{km}(s,t))$$

is given by

$$\hat{P}_k^{(\mathrm{LM})}(s,t) = \hat{p}^{(\mathrm{LM})}(s) \cdot \prod_{u \in (s,t]} (\mathbf{I} + \, \mathrm{d}\hat{\mathbf{A}}^{(\mathrm{LM})}(u)),$$

where $\hat{p}^{(\mathrm{LM})}(s)$ is a row vector with entry 1 for state k and 0 otherwise. The landmark state is k and $\hat{\mathbf{A}}^{(\mathrm{LM})}(u)$ is a matrix of Nelson-Aalen estimates, estimated from the landmark data set which selects subjects who are in state k at time s. Additionally assuming that $P(X(s)=k)>0$, the landmark Aalen-Johansen is a consistent estimator under the same assumptions as needed for the state occupation probabilities (Putter and Spitoni, 2016).

We emphasize that the landmarking approach, in general, uses less data than the standard Aalen-Johansen estimator. For illustration, consider an illness-death model without recovery, see also Section 3 below, but subject to right-censoring only. The Aalen-Johansen estimator of staying in the intermediate illness state on $[s,t]$ given illness at time s is a standard Kaplan-Meier-type estimator. This estimator will also use observed trajectories entering the illness state after time s, say, at time $\tilde{s}\in(s,t)$, and making a transition into the death state until time t. The landmarking approach will not use such trajectories. Now, also introduce left-truncation. The standard Aalen-Johansen estimator would use, say, a trajectory that enters the study at time $\tilde{s}$ in the illness state (and may even have been in the illness state at time s). But the landmarking approach, now extended to left-truncated data, will not use this trajectory. The difference to the situation without left-truncation is that landmarking would have used this trajectory, if it had been in the illness state at time s. However, even in the absence of left-truncation, landmarking would not have used the trajectory, if the patient had fallen ill after time s.

2.3. Resampling

For inference, resampling is arguably the most convenient approach, even in the time-inhomogeneous Markov case, see Bluhmki et al. (2018) for a recent in-depth discussion. For the non-Markov case, the methodological developments rely on a perfect i.i.d. setup with entirely unrelated left-truncation and right-censoring. Hence, it is natural to begin by exploiting the i.i.d. structure of the data and use Efron's bootstrap which draws with replacement from the units under study. For instance, to construct point-wise confidence intervals, consider $\hat{P}_{lm}^{(\mathrm{LM};*b)}(s,t)$, the b-th landmark Aalen-Johansen bootstrap estimate, $b=1,\dots,B$, where B is the number of bootstrap samples. Then the $(1-\alpha)$ bootstrap confidence interval for $P_{lm}^{(\mathrm{LM})}(s,t)$ is constructed using percentiles of the bootstrap distribution, i.e. the confidence interval is $(c^*(\alpha/2),c^*(1-\alpha/2))$, with $c^*(\alpha)$ the α-percentile of the bootstrapped $\hat{P}_{lm}^{(\mathrm{LM};*)}(s,t)$ (Davison and Hinkley, 1997, Chapter 5).

However, we may prefer the wild bootstrap in the time-inhomogeneous Markov case where non-random censoring may violate the i.i.d. assumption, and there is hope that it may also reasonably perform in the non-Markov i.i.d. setup. The idea is to consider the martingale decomposition (6) of the transition-specific counting processes. Re-expressing (6) on the level of individual increments,

$$\mathrm{d}M_{i;lm}(t)=\mathrm{d}N_{i;lm}(t)-Y_{i;l}(t)\cdot\alpha_{i;lm}(t|\mathcal{G}(t-))\,\mathrm{d}t,$$

we substitute these unknown martingale increments by $\mathrm{d}N_{i;lm}(t)$ times a standard normal random variable independent of the data. Generating a large number of the latter multipliers given the data, i.e., treating $\mathrm{d}N_{i;lm}(t)$ as fixed, is the basis of the wild bootstrap with the aim to mimick the asymptotic distribution of the transition-specific Nelson-Aalen estimators if used in (7) or in (16) of the Appendix, see the discussion below. Following the approach of Bluhmki et al. (2018) for the time-inhomogeneous Markov case, a transformation of such simulated martingale distributions along the compact derivative of the product integral yields another bootstrapping procedure.

Wild bootstrapping in the non-Markov setting is imperfect. In the Appendix, Equation (16) states

$$\begin{aligned}\hat{A}_{lm}(t)&=\int_0^t\frac{J_l(u)}{Y_l(u)}\,\mathrm{d}M_{lm}(u)+\int_0^t\frac{J_l(u)}{Y_l(u)}\sum_{i=1}^{n}Y_{i;l}(u)\cdot\alpha_{i;lm}(u|\mathcal{G}(u-))\,\mathrm{d}u\\&=:M_{lm}^{\star}(t)+A_{lm}^{\star}(t).\end{aligned}$$

The interpretation of Equation (16) is 'estimator equals error process plus approximately the quantity to be estimated'. Wild bootstrapping targets the error process and, hence, captures the variation in $M_{lm}^{\star}(t)$, but not in $A_{lm}^{\star}(t)$. In the Markov case, $\alpha_{i;lm}(u|\mathcal{G}(u-))$ equals the non-random $\alpha_{lm}(u)$, and $A_{lm}^{\star}(t)$ reduces to $\int_0^t J_l(u)\cdot\alpha_{lm}(u)\,\mathrm{d}u$ with $J_l=1$ asymptotically. The hope for the wild bootstrap to also perform reasonably in the non-Markov case is that $M_{lm}^{\star}$ may be the main driver of variation in many practical situations. To begin, this is motivated by the Markov case and should hold true if violations of the Markov assumption are only moderate or only affect some transitions. The latter idea is also pursued in Maltzahn et al. (2020) and related to the idea of Christiansen and Furrer (2021) to explain violations of the Markov assumption by a larger state space, then Markov although (in the words of the authors) possibly 'fictional'. Secondly, the variance of the Nelson-Aalen estimator is typically estimated at the jumps, building on the optional variation process of M_{lm} being N_{lm}, while the contribution of the last integral on the right hand side of the previous display would still be almost smooth. In a similar vein, Scheike (2002) found that martingale standard errors performed well, although the underlying model was a rate model for recurrent events, see also the discussion in Martinussen and Scheike (2007).

The motivation to consider wild bootstrapping is that in a time-inhomogeneous Markov setting as in Bluhmki et al. (2018), we could also allow for more general censoring schemes, not necessarily entirely random and violating the i.i.d. structure, as we will demonstrate in Section 3.2. There, we will consider non-random censoring caused by having observed a prespecified number of events, a common form of censoring in pharmaceutical trials, which also leads to dependent time-to-event trajectories. In contrast, drawing with replacement requires an i.i.d. structure in both non-Markov and Markov situations. We also note the resampling proposal by Glidden (2002) in a non-Markov setting which uses one standard normal multiplier per individual rather than per transition, see also the discussion in Bluhmki et al. (2019a).

3. Simulation and real data results

In both simulations and in the real data analysis, we focus on the illness-death model without recovery for two reasons: Firstly, the motivation from the real data analysis are hospital-acquired infections which will be represented by the intermediate 'illness'-state. Secondly, the illness-death model allows to concisely characterize departures from the Markov assumption. These occur, if the intensity of the illness-to-death transition also depends on the time of illness diagnosis. Section 3.1 uses simulations to study whether state occupation probabilities may be consistently estimated in a non-Markov model subject to random left-truncation. The simulation setting is based on a scenario used in Meira-Machado et al. (2006). Section 3.2 takes a closer look at using Efron's bootstrap or the wild bootstrap. For ease of presentation, we consider the Nelson-Aalen estimator of the illness-to-death transition — which 'captures' violations of the Markov assumption — and compare both bootstrapping procedures in a non-Markov setting and in a Markov setting. In the latter, censoring will not be random. Finally, simulations investigating the landmark Aalen-Johansen estimator are in Section 3.3 and the real data example in Section 3.4. The simulation settings for the non-Markov models in Section 3.2 and Section 3.3 are based on simulation settings used in Titman (2015) and Putter and Spitoni (2016). For evaluation of the Markov model in Section 3.2 we simulate an illness-death model with constant hazard rates. The number of bootstrap replications is 1000 for each simulation study.

3.1. State occupation probabilities

We present the results of a simulation study that assesses how well the Aalen-Johansen estimator for the state occupation probability does under random left-truncation.

The simulations are based on the multi-state simulation algorithm of Bluhmki et al. (2019b) and on a scenario used in Meira-Machado et al. (2006), who simulated an illness-death model without recovery with initial state 0, intermediate state 1 and absorbing state 2. Falling ill is modelled as a transition into state 1, death is modelled as a transition into state 2 and recoveries $1 \to 0$ are not modelled. The hazards of a $0 \to 1$ and $0 \to 2$ transitions are assumed to be constant and equal to 0.039 and 0.026, respectively. The waiting time in the initial state is generated using a constant hazard of 0.039 + 0.026. A binomial experiment then decides on whether the individual moves into state 1 with probability $0.039/(0.039 + 0.026)$.

For the individuals that move into state 1, two methods for generating times of arrival into state 2 are considered. The first simulation method, suggested by Couper and Pepe (2001), is to specify $Z_2 = (1 + d)Z_1$, where d is an arbitrary constant and Z_1 and Z_2 denote the time points of arrival in state 1 and 2, respectively. We set $d = 0.7$ in the following as in Allignol et al. (2014). Note, however, that fixing the departure time from the intermediate state given the arrival time violates both the Markov assumption and the assumption of continuous waiting times. The second simulation method avoids the latter drawback and uses a Cox model with time-dependent covariate to create the hazard function of a $1 \to 2$ transition. Let $\alpha_{12}(t|Z_1)$ be the hazard for a certain individual to move from state 1 to state 2, with $\alpha_0(t)$ the baseline hazard and β a constant coefficient. Then $\alpha_{12}(t|Z_1) = \alpha_0(t)\exp(\beta Z_1)$. The baseline hazard for the $1 \to 2$ transition is set to $\alpha_0(t) = \alpha_0 =$0.1 and the coefficient $\beta = 0.1$.

Random left-truncation times are sampled from a skew normal distribution (Azzalini, 1985) with location, scale and shape parameters chosen such that approximately 70% of the individuals are actually included in the study. Approximately 3% of all simulated individuals enter the study at time origin. The parameters are $(0, 10, 10)$ and $(0, 11.5, 10)$ for the "multiplication by a constant scenario" and the "Cox scenario", respectively.

We simulate 1000 studies with a sample size of 100 individuals. Figure 1 reports the average of the 1000 estimates of $P_{01}(0, t)$ as well as the simulation based 95% confidence intervals for the two scenarios. Also displayed is the true value $P_{01}(0, t)$ numerically approximated by computing the Aalen-Johansen estimator in a study without truncation with 100.000 individuals. As can be seen, the curves can almost not be distinguished, in line with the novel consistency result of Section 2 in a left-truncated scenario.

We also note that in a study of 100 independent individuals with, on average, about 3% present at time origin, we expect about $0.97^{100} \approx 4.8\%$ simulated studies with no one at risk at time origin. This has no effect on the initial distribution, because we are considering an illness-death model with one common initial state. For Nelson-Aalen and Aalen-Johansen estimates, this implies that one has to wait for individuals to enter the study before observed transitions can be counted. This does not affect the asymptotic consistency result, because there is a positive probability to be at-risk at time origin. However, in an actual study, it may motivate researchers to reconsider their starting time point.

3.2. Drawing with replacement from the data or wild bootstrapping?

As outlined above, departures from the Markov assumption in an illness-death model without recovery are captured by the transition intensity for moving from the intermediate into the absorbing state. For conciseness of presentation, we focus in the present subsection on the Nelson-Aalen estimator of the illness-to-death transition in our comparison of the relative merits of the two bootstrapping procedures at hand.

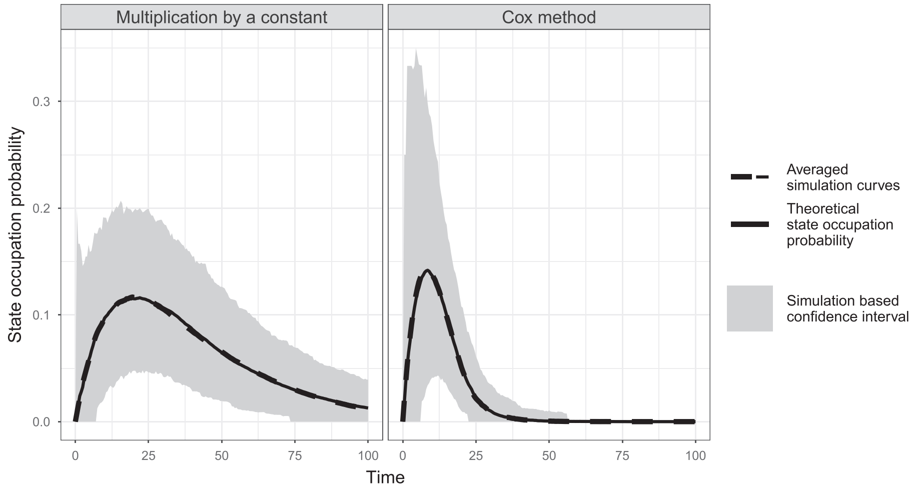

Fig. 1. Results of the simulation study for the scenario "multiplication by a constant" in the left panel and the "Cox method" on the right. The solid black line represents the true state occupation probability, the dashed line the averaged probabilities obtained from the simulated studies. The grey area is the simulation based 95% confidence interval.

Table 1
Comparison of empirical coverage probabilities of the 95% point-wise confidence intervals constructed using Efron's bootstrap and the wild bootstrap at different time-points (e.g. T_{15} := 15 units of time) in non-Markov illness death model

	Coverage (%)					
	Efron			Wild Bootstrap		
n	T_{15}	T_{20}	T_{25}	T_{15}	T_{20}	T_{25}
n=30	88	86	83	96	96	97
n=50	91	91	90	95	95	96
n=100	93	94	94	94	94	96

As a first step, we simulate data from an illness-death model without recovery as in the previous section. We introduce dependence between the waiting time in the initial state and the waiting time in the illness state by multiplying transition hazards by a common gamma frailty. Here, the frailty is a gamma-distributed random variable with mean and variance equal to 2. Moreover, we added exponentially distributed random right-censoring times. We chose the rate parameter equal to 1/30, which leads to an average censoring rate of 56%. We use the following constant transition hazards: $\alpha_{01} = 0.12$, $\alpha_{12} = 0.1$, $\alpha_{02} = 0.03$. We consider different sample sizes — 30, 50 and 100 individuals per study — and simulate 1000 studies for each scenario. Each study is evaluated at different time-points. For the construction of point-wise 95% confidence intervals, we use two different resampling techniques — Efron's bootstrap and the wild bootstrap technique. Both resampling methods are used to determine the $(1-\alpha/2)$ - quantiles of the studentized Nelson-Aalen estimator which are plugged-in in the standard asymptotic formula instead of using the quantiles of the standard normal distribution. We use the empirical variance of the bootstrapped Nelson-Aalen estimators as variance estimator. Table 1 compares the coverage probabilities of the 95% point-wise confidence intervals obtained from those two resampling methods for different sample sizes and at different time-points. Both methods provide coverage probabilities close to the nominal level of 95% for n=100. However, for the scenarios with a sample size of 30 or 50 individuals, the wild bootstrap approach still leads to coverage probabilities close to the nominal level, whereas the confidence intervals obtained by Efron's bootstrap were found to have smaller coverage probabilities. In summary, under an i.i.d. data structure both resampling approaches may provide reliable results in situations where the Markov property is in doubt. Our simulations indicate that for small sample sizes the wild bootstrap approach may perform better compared to Efron's bootstrap.

One major advantage of the wild bootstrap technique is that it is, in contrast to Efron's bootstrap, not limited to the strict situation with i.i.d. data structure. Thus, the wild bootstrap does not require random censoring. As the i.i.d. data structure is a requirement for the consistency of the Nelson-Aalen estimator in non-Markov models, we consider a Markov model subject to event-driven censoring, so-called type II censoring, to investigate the impact of the violation of the i.i.d. data assumption. In other words, the aim of the following simulation is to investigate possible advantages of wild bootstrapping when random censoring, but not the Markov property is in doubt. Type II censoring implies that all individuals will be censored at the time when a specified number of occurrences of the event of interest has taken place. Thus, type II censoring is not random censoring but it fulfils the independent censoring assumption of Andersen et al. (1993) and Aalen et al. (2008), in that it retains the form of the intensities of the counting processes as in (6).

Table 2
Comparison of empirical coverage probabilities of the 95% point-wise confidence intervals constructed using Efron's bootstrap and wild bootstrap at different time-points (e.g. $T_8 := 8$ units of time) in Markov models subject to type II censoring with censoring after m events have been observed.

		Coverage (%)					
		Efron			Wild Bootstrap		
n	m	T_8	T_{12}	T_{16}	T_8	T_{12}	T_{16}
n=50	m=25	46	62	72	65	86	94
n=80	m=40	68	80	85	82	93	95
n=100	m=50	75	85	87	88	94	95
n=200	m=100	90	91	93	94	94	94

Table 3
Average number of individuals under study ($\bar{n}$), bias, root mean square error (RMSE), and coverage (%). Cov E: Coverage Efron's bootstrap, Cov W: Coverage Wild Bootstrap. Trunc: The distribution of the left-truncation times, either uniform or exponential.

			AJ				LMAJ			
Trunc	m	$\bar{n}$	Bias	RMSE	Cov E	Cov W	Bias	RMSE	Cov E	Cov W
Simulation model 'Frailty'										
Unif	200	141	-0.0052	0.045	94	92	-0.0030	0.064	94	94
	500	354	-0.0047	0.029	95	93	0.0006	0.040	95	95
Exp	200	157	-0.0045	0.036	95	94	-0.0001	0.055	94	95
	500	392	-0.0049	0.023	94	93	0.0011	0.032	94	95
Simulation model 'non-Markov'										
Unif	200	153	-0.0280	0.057	90	88	0.0013	0.084	94	94
	500	381	-0.0291	0.043	85	83	0.0004	0.050	95	95
Exp	200	146	-0.0265	0.057	91	88	0.0021	0.077	93	94
	500	364	-0.0285	0.042	86	85	0.0013	0.047	95	97

In our simulation studies, we censor all individuals at the time when half of the individuals had an observed death event. We use constant hazard rates ($\alpha_{01} = 0.01$, $\alpha_{12} = 0.1$, $\alpha_{02} = 0.03$) and no staggered study entry. That means all individuals enter the study at time 0. This implies that all individuals within one study have the same censoring time. It has to be noted that the ill-to-death transition times are subject to internal left-truncation due to the $0 \rightarrow 1$ transition and we do not consider time points where most of the studies are already censored. Table 2 shows the coverage probabilities of the 95 % confidence intervals constructed using the two different resampling methods at different time points and for different sample sizes. It can be seen that in our simulation setting the wild bootstrap technique provides coverage probabilities closer to the nominal level compared to Efron's bootstrap for all considered scenarios. We conclude that there are scenarios where the wild bootstrap performs better than Efron's bootstrap.

3.3. The landmark Aalen-Johansen estimator

We now extend the simulation setting of Titman (2015) and Putter and Spitoni (2016). The data are simulated from an illness-death model without recovery. As in Titman (2015), we consider two processes, a non-Markov process induced via a common gamma frailty ('Frailty' model) and a pathological non-Markov process ('non-Markov' model). The baseline transition hazards are constant, with $\alpha_{01} = 0.12$, $\alpha_{02} = 0.03$ and $\alpha_{12} = 0.1$. For the 'Frailty' model, all hazards are multiplied by a common gamma frailty with mean and variance equal to 2. The frailty term introduces dependence between the waiting time until leaving the initial state of the illness-death model and the waiting time until absorption and, hence, a violation of the Markov assumption. In the 'non-Markov' scenario, $\alpha_{12}(t)$ is dependent on the state occupied at time 4, i.e.,

$$\alpha_{12}(t) = \begin{cases} 0.05, & \text{if } X(4) = 0, \\ 0.1, & \text{otherwise.} \end{cases}$$

Transition times are simulated as in Section 3.1. We consider sample sizes $m = 200$ and $m = 500$ and random left-truncation times following a uniform distribution with parameters $(-5, 28)$ and $(-1, 13)$ for the 'Frailty' and 'non-Markov' scenario, respectively. We consider also exponentially distributed left-truncation times with parameter 0.13. Here, nobody is starting in the study at time 0. Table 3 reports number of individuals simulated (m), average number of individuals in the study ($\bar{n}$, $\bar{n} < m$ because of left-truncation), bias, root mean squared error (RMSE), and the empirical coverage probabilities of the 95% bootstrap confidence intervals, for the Aalen-Johansen and landmark Aalen-Johansen estimates of $P_{01}(\tau_{0.15}, \tau_{0.45})$, where $\tau_{0.15}$ and $\tau_{0.45}$ correspond to the 15th and 45th percentiles of the time-to-absorption distribution whose values are taken from the supplementary material of Titman (2015). Note that the information on coverage will be more relevant for the unbiased estimator, but we have chosen to include the information for all estimators.

As in Putter and Spitoni (2016), the landmark Aalen-Johansen performs well. The Aalen-Johansen estimator is slightly more biased but displays the smaller RMSE. Regarding the landmark Aalen-Johansen estimator, both resampling methods provide confidence intervals with coverage probabilities close to the nominal 95% level, with a certain preference for the wild bootstrap. In contrast, the coverage probabilities of both resampling approaches may drop well below 90% for the Aalen-Johansen estimator.

3.4. The impact of MRSA infection among MRSA colonized patients on length of hospital stay

We consider data on patients colonised by Methicillin-resistant *staphylococcus aureus* (MRSA) from a prospective cohort study in 12 surgical units at the University of Geneva Hospitals, Switzerland, between July 2004 and May 2006 (De Angelis et al., 2011). MRSA carriage is not necessarily detected upon hospital admission, because a positive MRSA screening result may come in ‘delayed’ in the sense that the positive laboratory result becomes available or known only after admission. Hence, our population of interest are patients colonised by MRSA, the time scale of interest is time since hospital admission and the left-truncation time is the time of detecting MRSA in the screening process. Colonised patients who are discharged or die before a positive screening result becomes available are not included in the study.

MRSA colonization may lead to MRSA *infection* in hospital, and the more severe or potentially life-threatening MRSA infections are observed most frequently in healthcare settings. In this context and in the presence of limited financial resources, possible measures of infection control are weighed against the costs associated with hospital-acquired infection (Muto et al., 2003). To this end, excess length of stay following the infection is typically considered to be the main cost driver (Graves et al., 2010).

However, quantifying excess length of stay is complicated by the fact that hospital-acquired MRSA infection is a time-dependent exposure and a simple retrospective comparison of the distribution of length of stay of the infected with that of the only colonised must overestimate the prolonging effect of the infection as a consequence of ‘immortal time bias’ (Beyersmann et al., 2008; Suissa, 2008). We address this difficulty as follows: Firstly, we model occurrence of hospital-acquired MRSA infection as an intermediate state in an illness-death multi-state model, in which the initial state represents colonization, intermediate state 1 infection and absorbing state 2 discharge from the hospital. Secondly, we use landmarking to compare the residual length of stay (in terms of the transition probabilities) of those in the infectious state at the landmark with those still in the initial state of colonization.

We begin our analysis by checking the Markov assumption using a Cox model, estimate the proportion of currently infected and hospitalized patients using the Aalen-Johansen estimator of state occupation probabilities and finally compare the residual length-of-stay distributions between infected and only colonised for different landmarks using the landmark Aalen-Johansen estimator of transition probabilities. Recall that all analyses must account for the data being subject to left-truncation.

In order to check the Markov assumption we include the time of infection in a Cox proportional hazards model for the hazard of end-of-stay (Keiding and Gill, 1990). The hazard ratio (0.98, 95%-CI [0.97, 0.99]) is found to be significantly smaller than 1, indicating a departure from the Markov assumption.

Assuming random left-truncation times is reasonable since delays are most likely due to recording issues of positive molecular test results for MRSA. Furthermore, we checked the assumption of random left-truncation by fitting a Cox proportional hazard model on the marginal distributions of ‘time of hospital stay’ (T) and ‘time till infection or infection-free discharge (whatever came first)’ (T_0) including left-truncation times as covariates. The pair T and T_0 equivalently describes the multi-state process. We obtain a hazard ratio equal to 1 (95%-CI [0.99, 1.01]) for T and a hazard ratio equal to 1 (95%-CI [0.99, 1.01]) for T_0. This indicates that the assumption of random left-truncation times may be assumed to be fulfilled. It should be kept in mind that the assumption of random left-truncation can only be tested based on the observed data. If the truncation time is greater than the event time, both event time and truncation time remain unobserved. We refer to Tsai (1990) and Wellek (1990) for an in-depth discussion.

Figure 2 displays the estimated probability to be in the infectious state, i.e., the state occupation probability with pointwise 95% confidence intervals obtained by using wild bootstrap. We note that this probability is low.

Finally, Figure 3 provides a landmark analysis, displaying both landmark Aalen-Johansen estimates and the original Aalen-Johansen estimates of $P_{02}(s,t)$ and $P_{12}(s,t)$ for a selected range of landmark times $s \in \{2, 3, 4, 5, 8, 10, 15, 20, 25\}$, $s < t$. The prolonging effect of MRSA is illustrated by the fact that $\hat{P}_{02}(s,t)$ runs above $\hat{P}_{12}(s,t)$ for all s, though this effect is much more pronounced for s between 8 and 15 days. We also see that for these data, the Aalen-Johansen and landmark Aalen-Johansen estimators are close to each other in spite of the data being possibly non-Markov.

4. Discussion

Multi-state modelling is useful for investigating complex courses of a disease in a variety of medical disciplines, but usually comes with a time-inhomogeneous Markov assumption to facilitate the technical developments. In particular, estimating transition probabilities when the Markov assumption is in doubt has been an active field of research in recent years. To the best of our knowledge, one of the first proposals is due to Meira-Machado et al. (2006) who used Kaplan-Meier integrals for a randomly right-censored illness-death model. Their approach was simplified by Allignol et al. (2014) using competing risks techniques. Allignol et al. (2014) also used semi-parametric efficiency arguments to arrive at landmark transition probability estimators, also allowing for delayed study entry, see also Gerds et al. (2017). Their approach was then recently extended to arbitrary multi-state models by Titman (2015), see also de Uña-Álvarez and Meira-Machado (2015) for related

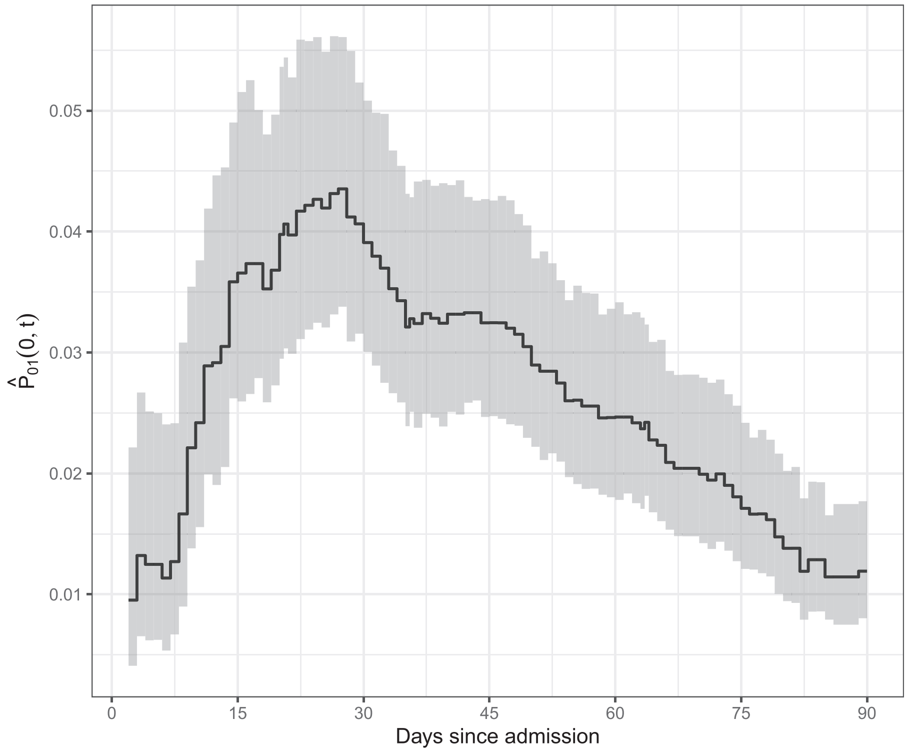

Fig. 2. Estimated state occupation probability $P_{01}(0,t)$ for the MRSA data along with 95% point-wise confidence interval based on 1000 bootstrap samples (wild bootstrap).

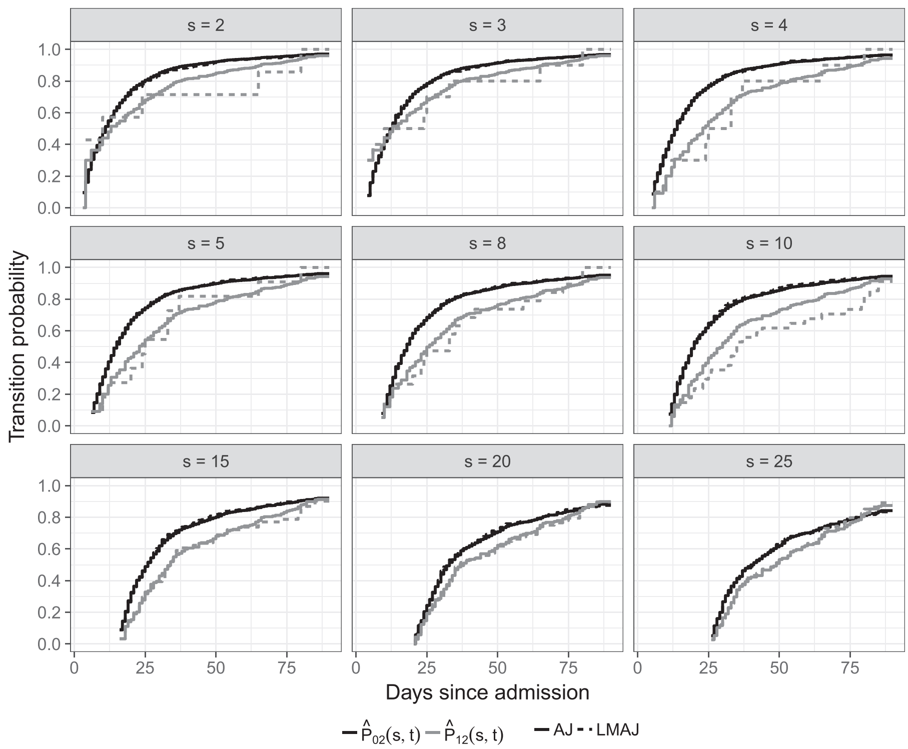

Fig. 3. Landmark analysis of the MRSA data. Each panel displays $P_{02}(s,t)$ (black) and $P_{12}(s,t)$ (grey) for $s \in \{2, 3, 4, 5, 8, 10, 15, 20, 25\}$ estimated by the Aalen-Johansen estimator (solid lines) and the landmark Aalen-Johansen estimator (dashed lines).

work. Arguably the most natural approach is due to Putter and Spitoni (2016) using a landmark Aalen-Johansen estimator and consistency of the Aalen-Johansen estimator for state occupation probabilities of non-Markov multi-state models subject to random right-censoring.

As clearly stated by Putter and Spitoni (2016), see their Appendix I, their approach crucially relies on consistency of the Aalen-Johansen estimator when estimating state occupation probabilities (Datta and Satten, 2001), but, as discussed above, the validity of the earlier result has been recently questioned by several authors. Our paper adds to the existing literature by establishing the consistency needed (see also Overgaard (2019) discussed above) and extending results to delayed study entry which is a common phenomenon in observational studies. Delayed study entry is naturally incorporated both into the counting process framework used in the present paper and in the landmarking approach and was therefore already considered in the (otherwise much more restricted) estimator by Allignol et al. (2014), see also the discussion in Gerds et al. (2017). Our proof also motivates use of the wild bootstrap resampling method which we further investigated in both non-Markov and Markov models using simulation studies. In contrast to Efron's bootstrap which draws with replacement from the data, the wild bootstrap is not limited to situations with i.i.d. data structure and, hence, could also be applied to censoring scenarios that are more complex than random censoring (Bluhmki et al., 2018), then, however, relying on the time-inhomogeneous Markov framework.

Returning to the landmarking approach, data subsetting to reflect the condition of a conditional probability is natural, but, as discussed earlier, results in using less data than in the Markov approach. This is already the case with data subject to right-censoring only, and delayed study entry reinforces this. Maltzahn et al. (2020) have recently attacked this problem using a 'hybrid' landmark Aalen-Johansen estimator, using transition-specific landmarking for those transitions only that are seen to cause the violation of the Markov assumption. Maltzahn et al. (2020) only consider right-censoring, but we expect that their arguments may be extended to left-truncation, combining their approach with the approach of this paper. Furthermore, de Uña-Álvarez and Mandel (2018) uses retrospective information before the truncation time to increase efficiency of their proposed estimators of transition probabilities in a general progressive multi-state model.

Declaration of Competing Interest

None declared.

Acknowledgments

We thank two anonymous reviewers whose constructive feedback substantially helped to improve our paper. Jan Beyersmann was partially supported by Grant BE 4500/4-1 of the German Research Foundation (DFG).

Appendix A. Proofs

Proof. of Theorem 2.1

To begin, note that we do not sample from P but from the conditional probability measure given study entry $Q(\cdot) = P(\cdot\,|Z)$. Dropping indices l as in $Y_{i;l}$ and lm as in N_{lm} denoting the transition type for ease of notation, we have that

$$\begin{aligned}\hat{A}(t) &= \int_0^t \frac{J(u)}{Y(u)}\,\mathrm{d}N(u)\\ &= \underbrace{\int_0^t \frac{J(u)}{Y(u)}\,\mathrm{d}M(u)}_{=M^\star(t)} + \underbrace{\int_0^t \frac{J(u)}{Y(u)}\sum_{i=1}^n Y_i(u)\cdot\alpha_i(u|\mathcal{G}(u-))\,\mathrm{d}u}_{=A^\star(t)}\end{aligned} \tag{16}$$

$M^\star(t)$ is a mean zero martingale with predictable variation process

$$V(t) = \int_0^t \frac{J(u)}{Y(u)^2}\sum_{i=1}^n Y_i(t)\cdot\alpha_i(u|\mathcal{G}(u-))\,\mathrm{d}u.$$

Because of Lenglart's inequality, we have that for any $\delta, \eta > 0$

$$Q\left(\sup_{[0,t]}|\hat{A} - A^\star| > \eta\right) \le \frac{\delta}{\eta^2} + Q(V(t) > \delta).$$

Assumption (9) implies

$$V(t) \le \int_0^t \frac{J(u)}{Y(u)}k(u)\,\mathrm{d}u$$

almost surely and it follows from (10) that

$$\sup_{[0,t]} |\hat{A} - A^\star| \xrightarrow{Q} 0.$$

To complete the proof, we still need to show

$$\sup_{[0,t]} \left| \int_0^s \left\{ \frac{J(u)}{Y(u)} \sum_{i=1}^n Y_i(u) \cdot \alpha_i(u|\mathcal{G}(u-)) \right\} - \alpha(u)\, \mathrm{d}u \right| \xrightarrow{Q} 0. \tag{17}$$

Under assumption (9) and using Gill's dominated convergence theorem (Andersen et al., 1993, Proposition II.5.3), it suffices to show point-wise convergence in probability of the integrand. We have

$$\frac{Y(u)}{n} \xrightarrow{Q} Q(Y_1(u) = 1) = Q(Y_i(u) = 1) \text{ for all } i = 1, \dots, n.$$

Next, note that the dependence of $\alpha_i(u|\mathcal{G}(u-))$ on the past of the observed data only constitutes dependence on ith observed past but not on that of other individuals j, $j \neq i$, because of independence across individuals. Using that left-truncation and right-censoring are random, we get that

$$Y_i(u) \cdot \alpha_i(u|\mathcal{G}(u-)), i = 1, \dots, n,$$

are i.i.d. random variables. Note that it does not suffice to require that the multi-state trajectories are i.i.d., but that we must additionally assume that our observational restrictions are entirely unrelated to the multi-state process in order to obtain that the random variables in the previous display are i.i.d.. Now, their average approaches the mean given by the following calculation

$$\begin{aligned} \mathrm{E}(Y_i(u) \cdot \alpha_i(u|\mathcal{G}(u-))\, \mathrm{d}u) &= \mathrm{E}\big(Q(\mathrm{d}N_i(u) = 1 \mid \mathcal{G}(u-))\big) \\ &= Q(\mathrm{d}N_i(u) = 1) \\ &= Q(Y_i(u) = 1) \cdot Q(\mathrm{d}N_i(u) = 1 \mid Y_i(u) = 1) \end{aligned}$$

and, recalling that we consider transitions $l \to m$,

$$\begin{aligned} &Q(\mathrm{d}N_i(u) = 1 \mid Y_i(u) = 1) \\ &\quad = P(X_i(u + \mathrm{d}u) = m, u + \mathrm{d}u \le C \mid X_i(u-) = l, L_i < u \le C_i) \\ &\quad = P(X_i(u + \mathrm{d}u) = m \mid X_i(u-) = l) = \alpha(u)\, \mathrm{d}u. \end{aligned}$$

In the previous display, the first equality holds, because $X_i(u-) = l, L_i < u$ implies study entry for a transient state l. Point-wise convergence of the integrand in (17) follows, also using (11), which completes the proof. Here, (11) is used to establish that not just the average over $Y_i(u) \cdot \alpha_i(u|\mathcal{G}(u-))$, $i = 1, \dots, n$, minus $\alpha(u)$ converges to zero, but also $J(u)$ times the average minus $\alpha(u)$. $\square$

The proof above has established Nelson-Aalen as a consistent estimator of the cumulative partly conditional transition rates in the presence of random left-truncation and right-censoring, using the common martingale reasoning of first demonstrating *convergence estimator minus compensator* followed by *convergence compensator minus target quantity*. This second step illustrates that the partly conditional transition rate has an interpretation as an average over individuals transition intensities, which has ties to an alternative proof based on an empirical process point of view. See Gill (1993, Chapter 6) on contrasting these two approaches for the Kaplan-Meier estimator, and Overgaard (2019) for the current multi-state setting. For the latter, the key idea is to consider the average number of $l \to m$ transitions and the average number at-risk in state l as estimators of their expectation and to view Nelson-Aalen as a continuous function of these empirical means; see also our discussion preceding Section 2.1 and, again, Overgaard (2019). We reiterate that our motivation to use martingale arguments has been, in brief, to correct the original argument, to motivate wild bootstrapping and to provide for an approach closer to the core of survival methodology, which is is illustrated by the relative ease with which left-truncation is handled and our contrasting of the approaches for event-driven censoring.

The result demonstrated above obviously also holds in the absence of both left-truncation and right-censoring, i.e., for complete data. The following proof now establishes consistent estimation of the state occupation probabilities, viewing Aalen-Johansen as a transformation of Nelson-Aalen. Because the transformation estimates the state occupation probabilities for complete data, it must estimate the same quantities when there is random left-truncation and right-censoring. This approach simplifies the inverse probability of censoring arguments of Datta and Satten (2001) and therefore naturally adapts to the inclusion of left-truncation, see also Glidden (2002).

Proof. of Theorem 2.2

The proof relies on the observation by Andersen et al. (1993, Section IV.4.1.4) for complete data that the entries of $\hat{p}(t)$ are the usual multinomial estimates, i.e., the number of trajectories observed in a specific state divided by n, if we chose $\hat{p}(0)$ as remarked after Theorem 2.2. In other words, Theorem 2.2 holds in the absence of both left-truncation and right-censoring.

Now, because product integration is a continuous functional (or operator) (Gill and Johansen, 1990), the assertion will follow as a consequence of the continuous mapping theorem, if the Nelson-Aalen estimator consistently estimates the same limit in the presence of random left-truncation and right-censoring as it does in the complete data case. This is precisely what Theorem 2.1 states, which completes the proof. Here, we view the multivariate Nelson-Aalen estimator as a random element of $D[0,t]^{K^2+K}$, $t \le \tau$, equipped with the max-supremum norm. □

References

Aalen, O., Borgan, O., Gjessing, H., 2008. Survival and event history analysis: a process point of view. Springer Science & Business Media.

Aalen, O., Johansen, S., 1978. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scandinavian Journal of Statistics 5, 141–150.

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

Andersen, P., Borgan, Ø., Gill, R., Keiding, N., 1993. Statistical Models Based on Counting Processes. Springer, New York.

Andersen, P., Keiding, N., 2002. Multi-state models for event history analysis. Statistical Methods in Medical Research 11 (2), 91–115.

Andersen, P.K., Angst, J., Ravn, H., 2019. Modeling marginal features in studies of recurrent events in the presence of a terminal event. Lifetime data analysis 25 (4), 681–695.

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics 12 (2), 171–178.

Beyersmann, J., Wolkewitz, M., Allignol, A., Grambauer, N., Schumacher, M., 2011. Application of multistate models in hospital epidemiology: advances and challenges. Biometrical Journal 53, 332–350.

Beyersmann, J., Wolkewitz, M., Schumacher, M., 2008. The impact of time-dependent bias in proportional hazards modelling. Statistics in Medicine 27, 6439–6454.

Bluhmki, T., Bramlage, P., Volk, M., Kaltheuner, M., Danne, T., Rathmann, W., Beyersmann, J., 2017. Time-to-event methodology improved statistical evaluation in register-based health services research. Journal of Clinical Epidemiology 82, 103–111.

Bluhmki, T., Dobler, D., Beyersmann, J., Pauly, M., 2019. The wild bootstrap for multivariate Nelson–Aalen estimators. Lifetime data analysis 25 (1), 97–127.

Bluhmki, T., Putter, H., Allignol, A., Beyersmann, J., on behalf of the COMBACTE-MAGNET consortium, 2019. Bootstrapping complex time-to-event data without individual patient data, with a view toward time-dependent exposures. Statistics in Medicine 38 (20), 3747–3763.

Bluhmki, T., Schmoor, C., Dobler, D., Pauly, M., Finke, J., Schumacher, M., Beyersmann, J., 2018. A wild bootstrap approach for the Aalen-Johansen estimator. Biometrics 74 (3), 977–985.

Christiansen, M.C., Furrer, C., 2021. Dynamics of state-wise prospective reserves in the presence of non-monotone information. Insurance: Mathematics and Economics 97, 81–98.

Cook, R.J., Lawless, J., 2007. The statistical analysis of recurrent events. Springer Science & Business Media.

Couper, D., Pepe, M.S., 2001. Modelling prevalence of a condition: chronic graft-versus-host disease after bone marrow transplantation. Statistics & Probability Letters 55 (4), 403–411.

Datta, S., Satten, G.A., 2001. Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov models. Statistics and Probability Letters 55 (4), 403–411.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap methods and their application. Cambridge university press.

De Angelis, G., Allignol, A., Murthy, A., Wolkewitz, M., Beyersmann, J., Safran, E., Schrenzel, J., Pittet, D., Harbarth, S., 2011. Multistate modelling to estimate the excess length of stay associated with meticillin-resistant staphylococcus aureus colonisation and infection in surgical patients. Journal of Hospital Infection 78, 86–91.

Dudley, R.M., Norvaiša, R., 2011. Concrete functional calculus. Springer.

Gasperoni, F., Ieva, F., Barbati, G., Scagnetto, A., Iorio, A., Sinagra, G., Di Lenarda, A., 2017. Multi-state modelling of heart failure care path: A population-based investigation from Italy. PLoS ONE 12 (6), e0179176.

Gerds, T., Beyersmann, J., Starkopf, L., Frank, S., van der Laan, M., Schumacher, M., 2017. From Statistics to Mathematical Finance Festschrift in Honour of Winfried Stute (Ed. Ferger, D et al.). Springer, New York, pp. 25–41. chapter The Kaplan-Meier integral in the presence of covariates: a review

Gill, R., 1993. Lectures on survival analysis, preprints 764 and 801, dept. Math., Univ. Utrecht.

Gill, R., Johansen, S., 1990. A survey of product-integration with a view towards application in survival analysis. Annals of Statistics 18 (4), 1501–1555.

Gillam, M.H., Ryan, P., Salter, A., Graves, S.E., 2012. Multi-state models and arthroplasty histories after unilateral total hip arthroplasties: introducing the summary notation for arthroplasty histories. Acta orthopaedica 83 (3), 220–226.

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

Graves, N., Harbarth, S., Beyersmann, J., Barnett, A., Halton, K., Cooper, B., 2010. Estimating the Cost of Health Care-Associated Infections: Mind Your p's and q's. Clinical Infectious Diseases 50 (7), 1017–1021.

Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 47 (260), 663–685.

Jepsen, P., Vilstrup, H., Andersen, P.K., 2015. The clinical course of cirrhosis: the importance of multistate models and competing risks analysis. Hepatology 62 (1), 292–302.

Keiding, N., Gill, R., 1990. Random truncation models and Markov processes. The Annals of Statistics 18, 582–602.

Maltzahn, N., Hoff, R., Aalen, O. O., Mehlum, I. S., Putter, H., Gran, J. M., 2021. A hybrid landmark Aalen-Johansen estimator for transition probabilities in partially non-Markov multi-state models. 1–24.

Martinussen, T., Scheike, T.H., 2007. Dynamic regression models for survival data. Springer Science & Business Media.

Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C., 2006. Nonparametric estimation of transition probabilities in a non-Markov illness-death model. Lifetime Data Analysis 12 (3), 325–344.

Müller, C., 2015. Nelson-Aalen and Aalen-Johansen estimators for randomly left-truncated and right-censored non-Markov multistate models with application to hospital epidemiology. Ulm University, Institute of Statistics Master's thesis.

Munoz-Price, L.S., Frencken, J.F., Tarima, S., Bonten, M., 2016. Handling time dependent variables: antibiotics and antibiotic resistance. Clinical Infectious Diseases 62 (12), 1558–1563.

Muto, C.A., Jernigan, J.A., Ostrowsky, B.E., Richet, H.M., Jarvis, W.R., Boyce, J.M., Farr, B.M., 2003. SHEA guideline for preventing nosocomial transmission of multidrug-resistant strains of staphylococcus aureus and enterococcus. Infection Control & Hospital Epidemiology 24 (5), 362–386.

Overgaard, M., 2019. State occupation probabilities in non-Markov models. Mathematical Methods of Statistics 28 (4), 279–290.

Pepe, M.S., Cai, J., 1993. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. Journal of the American Statistical Association 88, 811–820.

Putter, H., Spitoni, C., 2016. 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.

Scheike, T.H., 2002. The additive nonparametric and semiparametric Aalen model as the rate function for a counting process. Lifetime Data Analysis 8 (3), 247–262.

Schmoor, C., Schumacher, M., Finke, J., Beyersmann, J., 2013. Competing risks and multistate models. Clinical Cancer Research 12, 12–21.

Slama, R., Ballester, F., Casas, M., Cordier, S., Eggesbø, M., Iniguez, C., Nieuwenhuijsen, M., Philippat, C., Rey, S., Vandentorren, S., Vrijheid, M., 2014. Epidemiologic tools to study the influence of environmental factors on fecundity and pregnancy-related outcomes. Epidemiologic Reviews 36 (1), 148–164.

Suissa, S., 2008. Immortal time bias in pharmacoepidemiology. American Journal of Epidemiology 167 (4), 492–499.
Titman, A.C., 2015. Transition probability estimates for non-Markov multi-state models. Biometrics 71 (4), 1034–1041.
Tsai, W.-Y., 1990. Testing the assumption of independence of truncation time and failure time. Biometrika 77 (1), 169–177.
de Uña-Álvarez, J., Mandel, M., 2018. Nonparametric estimation of transition probabilities for a general progressive multi-state model under cross-sectional sampling. Biometrics 74 (4), 1203–1212.
de Uña-Álvarez, J., Meira-Machado, L., 2015. Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study. Biometrics 71 (2), 364–375.
Wellek, S., 1990. A nonparametric model for product-limit estimation under right censoring and left truncation. Stochastic Models 6 (4), 561–592.