

Integrating relative survival in multi-state models—a non-parametric approach

Damjan Manevski¹ , Hein Putter² , Maja Pohar Perme¹,
Edouard F Bonneville² , Johannes Schetelig³,
and Liesbeth C de Wreede²

Statistical Methods in Medical Research

2022, Vol. 31(6) 997–1012

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09622802221074156

journals.sagepub.com/home/smm

Abstract

Multi-state models provide an extension of the usual survival/event-history analysis setting. In the medical domain, multi-state models give the possibility of further investigating intermediate events such as relapse and remission. In this work, a further extension is proposed using relative survival, where mortality due to population causes (i.e. non-disease-related mortality) is evaluated. The objective is to split all mortality in disease and non-disease-related mortality, with and without intermediate events, in datasets where cause of death is not recorded or is uncertain. To this end, population mortality tables are integrated into the estimation process, while using the basic relative survival idea that the overall mortality hazard can be written as a sum of a population and an excess part. Hence, we propose an upgraded non-parametric approach to estimation, where population mortality is taken into account. Precise definitions and suitable estimators are given for both the transition hazards and probabilities. Variance estimating techniques and confidence intervals are introduced and the behaviour of the new method is investigated through simulations. The newly developed methodology is illustrated by the analysis of a cohort of patients followed after an allogeneic hematopoietic stem cell transplantation. The work has been implemented in the R package `mstate`.

Keywords

Multi-state model, relative survival, mortality tables, competing risks, `mstate`

1 Introduction

Multi-state models provide a framework for simultaneously analyzing competing events and sequences of events. In the medical field, these models help to study the impact of intermediate events on the prognosis of patients, and allow to estimate separate probabilities of death with and without the intermediate event at multiple time horizons. Interest can also be in distinguishing between deaths due to the studied disease and its treatment (excess mortality) and deaths due to other (population) causes. This is of particular interest in an older patient population where the risk of dying due to other causes is high and may considerably contribute to the overall proportion of deaths. However, cause of death is often not reported,¹ unreliable^{2,3} or cannot unequivocally be attributed to the disease or other causes.

Our motivating example comes from the study of outcomes for patients after an allogeneic hematopoietic stem cell transplantation (alloHCT) for myelodysplastic syndromes (MDSs) or secondary acute myeloid leukemia (sAML).⁴ Patients with these indications represent the oldest major patient population referred to alloHCT with a median age of 58 years at alloHCT (registry data from 2012). These patients are at significant risk for two competing failures: relapse of the underlying disease and non-relapse mortality (NRM), which for a large part is due to the transplantation and pre-treatment. The occurrence of

¹Institute for Biostatistics and Medical Informatics, Faculty of Medicine, University of Ljubljana, Slovenia

²Department of Biomedical Data Sciences, Leiden University Medical Center, the Netherlands

³Medical Clinic I, University Hospital, Germany

Corresponding author:

Damjan Manevski, Institute for Biostatistics and Medical Informatics, Faculty of Medicine, University of Ljubljana, Vrazov trg 2, 1000 Ljubljana, Slovenia.

Email: damjan.manevski@mf.uni-lj.si

relapse leads to a very poor prognosis. We investigated the contribution of population mortality to both death after relapse (DaR) and NRM. This enabled us to estimate the probability of excess NRM, which especially for older patients and for long-term outcomes may provide a better estimate for treatment-related mortality than all NRM. The current paper discusses and extends the model introduced in that paper.

When reliable information on causes of death is not available in the medical data, one needs to address this issue using external data. This is the core idea of the relative survival field⁵: the observed data is merged with general population mortality tables⁶ to indirectly enable the estimation of cause-specific hazards and probabilities.^{7,8} In this work, we extend the ideas of the relative survival methodology to a Markov time-inhomogeneous multi-state model where we focus on non-parametric estimation of transition hazards and probabilities. We provide a theoretical foundation, a study of behaviour of the methods under different scenarios and a software implementation in R that makes the methodology readily available to other users.

These measures have also been the focus of the relative survival field, which has seen considerable development in the last decade.^{5,9,10} With the generalisation to multi-state models, a new level of complexity is added. The main contribution of this paper is to discuss which measures may be of main interest in this setting and how to estimate them. We propose estimators for cumulative hazards and transition probabilities and study their properties, including their (asymptotic) variances. To the best of our knowledge, this is the first extension of relative survival to non-parametric estimation in a Markov multi-state model, whereas other semi-parametric^{11–13} or parametric^{14,15} approaches have already been proposed. An extensive simulation study of properties of the estimators in terms of bias, standard error and coverage probability for confidence intervals (CIs) has thus been performed. The simulation study is also the first (as to our knowledge) to extensively investigate the behaviour of different confidence interval methods for non-parametric multi-state models.

The article is structured as follows: in Section 2, we present the main theoretical grounds of our work. In Section 3, we present simulation results by which this new approach is evaluated. In Section 4, we will reanalyze the motivating dataset to illustrate the behaviour and interpretation of the new model. This also serves as a step-by-step overview of how such an analysis can be performed in R using the newly developed R code. In Section 5, we discuss the main findings of the article and provide conclusions.

2 Extended multi-state model integrating relative survival

We start by introducing the theoretical background of multi-state models and relative survival needed for our work in Sections 2.1 and 2.2. We then present our proposed extension in Section 2.3.

2.1 General multi-state model

A multi-state model is a stochastic process $(X(t), t \in [0, \tau))$ with a finite state space S of size K , $\tau \in \mathbb{R}^+$, where the sample paths are assumed to be right-continuous, i.e., $X(t+) = X(t)$. Thus, $X(t)$ represents the state at which the process (or a given individual) is present at time t . A transition from state h to state j takes place when an individual at risk in state h experiences an event that makes them enter state j .

Transition probabilities are defined in the following way

$$P_{hj}(s, t) = \text{Prob}(X(t) = j | X(s) = h, F_s),$$

for every combination of states $h, j \in S$; $s, t \in [0, \tau)$, $s \leq t$, and F_s containing the complete history (i.e. filtration) up to time s . In our work, we will assume that $X(t)$ is a Markov process so that the transition probability definition simplifies to

$$P_{hj}(s, t) = \text{Prob}(X(t) = j | X(s) = h). \quad (1)$$

A Markov model implies a clock-forward approach, that is, time is always measured since the same starting time, usually entry in the starting state.

From this, we can define transition hazards (or transition intensities) as the derivative of $P_{hj}(s, t)$ with respect to t (which we assume to exist), evaluated at $s = t$

$$\lambda_{hj}(t) = \lim_{\Delta t \rightarrow 0} \frac{P_{hj}(t, t + \Delta t) - P_{hj}(t, t)}{\Delta t};$$

the cumulative hazard is then defined as

$$\Lambda_{hj}(t) = \int_0^t \lambda_{hj}(u) \, du.$$

Apart from the definition in formula (1), transition probabilities can also be defined in a matrix form using a product integral

$$\mathbf{P}(s, t) = \prod_{(s,t]} (\mathbf{I} + \lambda(u)du). \tag{2}$$

where $\mathbf{P}(s, t)$ represents a square matrix of size $K \times K$. In such a matrix, every entry represents the transition probability of being in a given state at time t conditional on being in a given state at time s . Similarly, $\lambda(t)$ is a matrix of size $K \times K$ containing the hazards $\lambda_{hj}(t)$ for all possible transitions $h \rightarrow j$, $h \neq j$; whereas diagonal values are defined as $\lambda_{hh}(t) = -\sum_{j \neq h} \lambda_{hj}(t)$. Note that by $\Lambda(t)$ we denote the cumulative hazard equivalent of $\lambda(t)$.

Both transition hazards and probabilities are of key interest for our work. In Figure 1, we introduce the multi-state model which will be used as the leading example in the article. In this model, all patients with a certain disease are alive at time zero and have no change of their underlying disease status (e.g. relapse). From the starting state they can then experience two competing risks: relapse (an intermediate event) or death (i.e. non-relapse mortality, NRM; an absorbing state). Death can also occur after relapse. We call this state death after relapse (DaR) which is also an absorbing state. Thus, the model in Figure 1 is the usual illness-death model where the death state is split into two different states in order to distinguish the probabilities of dying before or after relapse (NRM and DaR, respectively). The two probabilities sum up to the total probability of dying. Although the intermediate event considered is relapse, any other relevant event could be considered, for example, recurrence, progression, adverse event, additional treatment, or recovery. In the following, we assume that each individual has the same censoring time for the different events for which they are at risk (which will typically differ between individuals), that all relapses before the individual censoring times are recorded in the dataset, and that the event times are observed exactly.

This model contains the key features of a multi-state model: combinations of competing risks and series of events. All theory and the implementation in software discussed below are also valid when the model is extended with additional states. They also hold when individuals start in different states or are only observed after a certain timepoint (left-truncated observations).

We now turn to estimation. The Nelson-Aalen estimator is the common non-parametric estimator used for the cumulative transition hazard

$$\widehat{\Lambda}_{hj}(t) = \int_0^t \frac{dN_{hj}(u)}{Y_h(u)}, \tag{3}$$

where $N_{hj}(u)$ represents the counting process that gives the number of $h \rightarrow j$ transitions in the time interval $[0, u]$, whereas $Y_h(u) = \sum_{i=1}^n Y_{h,i}(u)$ gives the size of the at-risk set, that is, the number of individuals present in state h just before time u ($Y_{h,i}(u)$ indicates whether the i th individual is present in state h just before time u).

For estimating the variance of the Nelson-Aalen estimator the Greenwood estimator is used, as commonly suggested in the literature^{16–19}:

$$\widehat{\text{Var}}_G(\widehat{\Lambda}_{hj}(t)) = \int_0^t \frac{dN_{hj}(u)}{Y_h(u)(Y_h(u) - dN_{hj}(u))}. \tag{4}$$

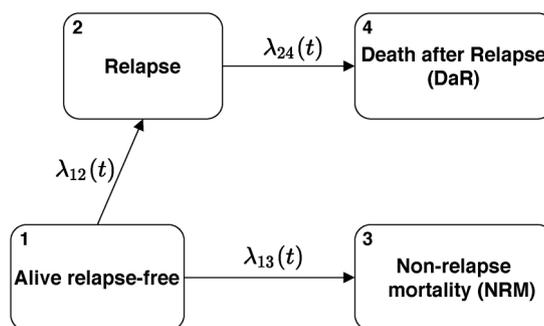


Figure 1. Transition diagram for the basic multi-state model.

For estimating transition probabilities the Aalen-Johansen estimator is standardly used; it is obtained from formula (2) by replacing $\Lambda(t)$ with the corresponding estimator $\hat{\Lambda}(t)$:

$$\hat{\mathbf{P}}(s, t) = \prod_{(s,t]} \mathbf{I} + d\hat{\Lambda}(u), \tag{5}$$

or written in discrete form: $\hat{\mathbf{P}}(s, t) = \prod_{u \in (s,t]} (\mathbf{I} + \Delta\hat{\Lambda}(u))$, where u indicates all event times in $(s, t]$ in the dataset.

We point out that both the Nelson-Aalen and Aalen-Johansen estimator are consistent when the multi-state model is Markov. For a non-Markov process, the Nelson-Aalen estimator is also consistent.²⁰ We also note that by using the landmark Aalen-Johansen estimator, each $P_{hj}(s, t)$ can be estimated consistently.²¹

For estimating the variance of transition probabilities we can again use the Greenwood estimator. We denote it as $\widehat{\text{Var}}_G(\hat{\mathbf{P}}(s, t))$, for which an exact definition is available in previous literature.^{22,16}

Here we have provided only the basic tools from multi-state models that are needed for this work; a more thorough and theoretical introduction can be found in previous work.^{23–25}

2.2 Relative survival

We will now introduce the main ideas of relative survival which are relevant for this work, using the basic survival setting where only one event (death) is reported in the data. Suppose we are interested in distinguishing between two causes of death—death due to the disease and its treatment and due to other (population) causes. To make this distinction we use population mortality tables as an additional source of information. Such mortality tables give hazards for every individual based on demographic covariates which have to be present both in the dataset and in the mortality tables (e.g. age, sex, year, country) to allow for proper matching. This enables us to evaluate the amount of deaths due to other (population) causes in the diseased population.

This information has to be appropriately incorporated in a model. We assume that the (observed) hazard of dying $\lambda(t)$ is the sum of two (unobserved) hazards, the hazard of dying because of the disease (i.e. the excess hazard, denoted as $\lambda_E(t)$) and the hazard of dying because of other causes (the population hazard, denoted as $\lambda_P(t)$). This assumption characterizes the additive model

$$\lambda(t) = \lambda_E(t) + \lambda_P(t). \tag{6}$$

The model is illustrated in Figure 2. In the interest of clarity we give an exact definition of these hazards. Let T_E be the time to excess death and T_P time to population death (i.e. death from other causes). We denote the minimum of these two times as T , the time of death. Time T can be also subject to a censoring time T_C which we assume to be non-informative. Hence, $\tilde{T} = \min(T, T_C)$ is observed, together with the event indicator at \tilde{T} (event or censoring). Then the hazard definition follows

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}; \lambda_E(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T_E \leq t + \Delta t | T > t)}{\Delta t}; \lambda_P(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T_P \leq t + \Delta t | T > t)}{\Delta t}.$$

The two hazards $\lambda_E(t)$ and $\lambda_P(t)$ are referred to as cause-specific hazards²⁶ in analogy to the standard competing risks model where both causes of death are observed.

It is important to note that in the definition of $\lambda_E(t)$ and $\lambda_P(t)$ we condition on $\{T > t\}$. This means that we condition on the observed time and “stick to the real world”.²⁷ Another possibility would be to condition on $\{T_E > t\}$ and $\{T_P > t\}$, respectively, which is commonly done when defining net survival.⁵ A more in-depth study of hazard definitions has been previously provided.^{10,5,9} We further assume that the population of interest is small compared to the general population, that is, that removing the subpopulation of interest from population life tables would have negligible effect on λ_P .



Figure 2. Illustration of the additive model given in formula (6).

The cumulative hazard $\Lambda(t) = \int_0^t \lambda(u)du$ can be estimated using the Nelson-Aalen estimator. The population hazard $\lambda_P(t)$ is estimated using mortality tables: for the i th individual we can obtain $\lambda_{P_i}(t)$ based on demographic covariate values. An estimator for the cumulative version of $\lambda_P(t)$ is then defined as

$$\widehat{\Lambda}_P(t) = \int_0^t \frac{\sum_{i=1}^n Y_i(u) d\Lambda_{P_i}(u)}{Y(u)}, \tag{7}$$

based on n individuals in the dataset. The definition of $\widehat{\Lambda}_P(t)$ corresponds to the definition of $\Lambda_P(t)$ when written with respect to the covariate distribution, see formula (1) in the Supplemental Material where we have also provided a derivation.

The estimator for the cumulative excess hazard $\Lambda_E(t)$ is defined as the difference of the estimators for the observed and population cumulative hazards, that is,

$$\widehat{\Lambda}_E(t) = \int_0^t \frac{dN(u)}{Y(u)} - \int_0^t \frac{\sum_{i=1}^n Y_i(u) d\Lambda_{P_i}(u)}{Y(u)}, \tag{8}$$

where $N(u)$ represents the number of events in $[0, u]$.

2.3 Extended multi-state model integrating relative survival

2.3.1 Estimation of hazards and probabilities

We will now extend this basic multi-state model presented in Section 2.1 with the relative survival approach introduced in Section 2.2. Thus, we will split death states in a multi-state model into excess and population counterparts. To better illustrate the idea, we will extend the example given in Figure 1. By matching the observed data with population mortality data, we would like to estimate the amount of population and excess deaths, both before and after the intermediate event. Such a model is illustrated in Figure 3—although the exact cause of death is not present in the dataset (the observed data are the same as before), we show two arrows and two additional death states by which we differentiate between excess and population death.

We now discuss the quantities of interest and their estimation under this model. We start with the hazards. For transitions that do not reach death (e.g. the transition from Alive relapse-free to Relapse, see Figure 3), the hazard is the same as in the standard multi-state model (Figure 1). It can again be estimated by the Nelson-Aalen estimator (3). On the other hand, each transition reaching death is now split into two (compare transitions 2 and 3 in Figure 1 to the split transitions in Figure 3). The hazards for these transitions have the same form as in Section 2.2. The estimators follow from formulas (7) and (8) which we now write in general form. For a transition reaching death $h \rightarrow j$ the estimators of the cumulative hazards for the population and excess transitions, denoted by (hj, P) and (hj, E) , are

$$\widehat{\Lambda}_{hj,P}(t) = \int_0^t \frac{\sum_{i=1}^n Y_{h,i}(u) d\Lambda_{P_i}(u)}{Y_h(u)}; \quad \widehat{\Lambda}_{hj,E}(t) = \int_0^t \frac{dN_{hj}(u)}{Y_h(u)} - \int_0^t \frac{\sum_{i=1}^n Y_{h,i}(u) d\Lambda_{P_i}(u)}{Y_h(u)}. \tag{9}$$

The estimators provided in formula (9) have the same form as in formulas (7), (8), the difference being that now through the at-risk process $Y_h(t)$ and counting process $N_{hj}(t)$ we can take into account additional competing risks (e.g. relapse) and left-truncation. Transitions from the intermediate state to death are usually left-truncated since individuals experience the transition to the intermediate event at different timepoints after time 0. Partly as a consequence of the later entry in the state and partly because some demographic features might be associated with an increased risk for the intermediate event,

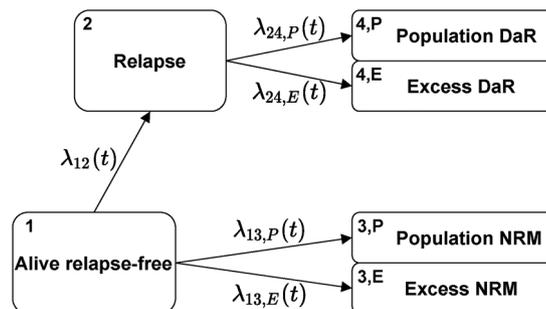


Figure 3. Transition diagram for the extended multi-state model.

demographic covariate values of those who experienced the intermediate event tend to be different from the whole group at baseline. These aspects are properly taken into account by continuously updating the population hazard based on the characteristics of the individuals observed to be at risk for each transition at each timepoint.

We also examine some additional characteristics of the estimators $\widehat{\Lambda}_{hj,P}(t)$ and $\widehat{\Lambda}_{hj,E}(t)$. First, we point out that the cumulative hazard $\Lambda_{hj}(t)$ is continuous whereas $\lambda_{hj}(t)$ may or may not be continuous. Compared to the Nelson-Aalen estimator $\widehat{\Lambda}_{hj}(t) = \int_0^t \frac{dN_{hj}(u)}{Y_i(u)}$ which is a step function (jumps occur only at event times), $\widehat{\Lambda}_{hj,P}(t)$ as defined in (9) is continuous and piecewise linear, since $\lambda_{P_i}(u)$ is taken from mortality tables. Consequently, $\widehat{\Lambda}_{hj,E}(t)$ is defined as a difference of a step and a continuous function (and is càdlàg, like $\widehat{\Lambda}_{hj}(t)$). In practice, the estimation is done on small intervals (usually on a daily basis). However, $\widehat{\Lambda}_{hj,P}(t)$ and $\widehat{\Lambda}_{hj,E}(t)$ are then evaluated at event times only (like the Nelson-Aalen estimator $\widehat{\Lambda}_{hj}(t)$) and they are taken to be constant between event times. This procedure causes some bias, which is mostly negligible. The same approach is taken in Pavlič and Pohar Perme²⁸ where it is also further discussed (see Section 3.4). Nonetheless, our R implementation allows for additional evaluation of the estimators at any choice of timepoints (e.g. on a daily basis). An illustration of the estimation process is given in the Supplemental Material, Section S2.

Having defined the estimates for transition hazards, we now turn to transition probabilities. Estimator $\widehat{\mathbf{P}}(s, t)$ is again defined in matrix form (5), whereby $\lambda(t)$ is extended using excess and population hazards for all split transitions. The product integral (5) then provides estimates, where for every split transition $h \rightarrow j$ we obtain estimates $\widehat{P}_{hj,P}(s, t)$ and $\widehat{P}_{hj,E}(s, t)$, whereas for all remaining transitions the definition of the estimator of the transition probabilities stays the same. In practice, transition probabilities are also estimated on small intervals but evaluated at event times only—a corresponding example is provided in the Supplemental Material. We further note that in the commonly used setting where all patients begin in the starting state at time zero and only transition probabilities conditioning on this point of departure are considered, transition probabilities are equivalent to state occupation probabilities.²⁴

There are also other measures that are common in relative survival. One such measure is net survival.⁵ However, it is questionable whether this measure is of interest for a multi-state model as its interpretation becomes less straightforward.

2.3.2 Variance estimation

We consider two approaches for estimating the variances of hazard and probability estimators—one based on the Greenwood estimator and the second one based on non-parametric bootstrap (simple resampling with replacement of individuals).

We start with transition hazards: for non-split transitions we use the usual Greenwood estimator (4) or bootstrap. To estimate the variances of the split hazards $\widehat{\Lambda}_{hj,E}(t)$ and $\widehat{\Lambda}_{hj,P}(t)$ we also consider two options. In the first, we assume that $\text{Var}(\widehat{\Lambda}_{hj}(t)) = \text{Var}(\widehat{\Lambda}_{hj,E}(t))$, thus we take $\widehat{\text{Var}}_G(\widehat{\Lambda}_{hj}(t))$ as the estimator of $\text{Var}(\widehat{\Lambda}_{hj,E}(t))$. In this case, one assumes that the population hazard $\Lambda_{hj,P}(t)$ is fixed and has variance zero. This assumption is often made because of the use of mortality tables which are deterministic. However, as is clear from formula (9), the estimator $\widehat{\Lambda}_{hj,P}(t)$ is not deterministic since $Y_h(t)$ brings some variability through the composition of the study population in terms of the demographic covariates (age, sex, year, etc.). As a second option for evaluating variability we use the bootstrap, which will help investigate if this additional assumption is reasonable.

Analogously, the variance of estimates for the transition probabilities is estimated using the Greenwood and bootstrap options.

2.3.3 CI methods

We consider CI methods for transition hazards and probabilities. Although the model of interest is still the model introduced in the previous subsection, the methods are general and can be applied in any non-parametric multi-state Markov model. We denote by θ the transition hazard or probability at a given time for a given transition and $\hat{\theta}$ its corresponding estimate. By denoting $\widehat{\text{Var}}_G(\hat{\theta})$ and $\widehat{\text{Var}}_{boot}(\hat{\theta})$ to be the Greenwood and bootstrap estimators of the variance, respectively, we define the following CI methods.

(CI1) *Plain scale using Greenwood (plain.G)*: a symmetrical $(1 - \alpha)\%$ CI on the plain scale using the Greenwood estimator for the variance

$$\hat{\theta} \pm \Phi\left(1 - \frac{\alpha}{2}\right) \cdot \sqrt{\widehat{\text{Var}}_G(\hat{\theta})}, \quad (10)$$

where Φ denotes the cdf of the standard normal distribution.

(CI2) *Plain scale using bootstrap (plain.boot)*: a symmetrical $(1 - \alpha)\%$ CI on the plain scale using the bootstrap estimator for the variance

$$\hat{\theta} \pm \Phi\left(1 - \frac{\alpha}{2}\right) \cdot \sqrt{\widehat{\text{Var}}_{boot}(\hat{\theta})}. \tag{11}$$

(CI3) *Log scale using bootstrap (log.boot)*: a $(1 - \alpha)\%$ CI on the log scale (based on the delta method) using the bootstrap estimator for the variance

$$\hat{\theta} \cdot \exp\left\{\pm \Phi\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sqrt{\widehat{\text{Var}}_{boot}(\hat{\theta})}}{\hat{\theta}}\right\}. \tag{12}$$

(CI4) *Quantiles using bootstrap (q.boot)*: a $(1 - \alpha)\%$ CI where the limits of the CI are taken to be the 2.5% and 97.5% quantiles of $\hat{\theta}$ over the bootstrap samples.

We define two additional methods which we use for comparison with the previous four methods:

(CI5) *Logit scale using bootstrap (logit.boot)*: a $(1 - \alpha)\%$ CI on the logit scale using the bootstrap estimator for the variance

$$\left[\frac{\hat{\theta}_l}{1 + e^{\hat{\theta}_l}}, \frac{\hat{\theta}_u}{1 + e^{\hat{\theta}_u}} \right], \quad \text{where } \hat{\theta}_{l/u} = \log\left(\frac{\hat{\theta}}{1 - \hat{\theta}}\right) \mp \Phi\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\widehat{\text{Var}}_{boot}(\hat{\theta})}{\hat{\theta}(1 - \hat{\theta})}}.$$

(CI6) *Complementary log-log using bootstrap (cloglog.boot)*: a $(1 - \alpha)\%$ complementary log-log CI using the bootstrap estimator for the variance

$$\left[\exp\left(-\exp\left\{\log(-\log\hat{\theta}) - \Phi\left(1 - \frac{\alpha}{2}\right) \frac{\sqrt{\widehat{\text{Var}}_{boot}(\hat{\theta})}}{\hat{\theta} \log\hat{\theta}}\right\}\right), \exp\left(-\exp\left\{\log(-\log\hat{\theta}) + \Phi\left(1 - \frac{\alpha}{2}\right) \frac{\sqrt{\widehat{\text{Var}}_{boot}(\hat{\theta})}}{\hat{\theta} \log\hat{\theta}}\right\}\right) \right].$$

3 Simulations

We analyse the characteristics of the developed methodology presented in Section 2.3 through simulations. First, in Section 3.1, we introduce all details regarding how the simulations were performed and in Section 3.2, we present the results. All simulations were done in R, version 3.5.2²⁹ using the packages `mstate`³⁰ and `relsurv`.²⁸ The main simulation functions have been added in a separate .R-file (`simulation_functions.R`) in the Supplemental Material.

3.1 Defining the simulation study

In this section, we present the aims, data-generating mechanisms, estimands, methods, and performance measures (i.e. ADEMP) used for simulation. These are reported as proposed in the tutorial by Morris et al.³¹

3.1.1 Aims

Our main goal is to investigate how the newly proposed estimators for the extension of multi-state models with population mortality perform in a wide range of simulation scenarios. We will evaluate their bias, standard errors (SE) and coverage probabilities (CPs).

3.1.2 Data-generating mechanisms

Transition diagram: we use the same transition diagram as shown in Figure 3. At time 0, all individuals are alive event-free. All individuals are followed up to a maximum of 10 years.

Generating event times: for every individual we generate a trajectory, where the variables and event times are obtained in the following way:

1. Demographic variables D are simulated independently for every individual: age (uniform), sex (binary), year (uniform). In all simulations, sex is distributed Bernoulli(0.5) and year of diagnosis is taken Uniform[1990-01-01, 2000-01-01].

- Age is also simulated with the uniform distribution on a prespecified interval that will be adjusted for every simulating scenario so that the amount of population deaths is controlled.
2. We first have three event times for the following competing risks: relapse, population NRM, and excess NRM. In practice, we only observe NRM and relapse, but in the simulations we also generate the underlying population and excess times to death. If an individual reaches state relapse, times to population DaR and excess DaR are then simulated conditional on the time of relapse. All event times are simulated independently.³²
 3. Times to death from population causes are simulated using population mortality tables for Slovenia (which are available in R-package `reلسurv`, object `slopop`). For simulating the transition from relapse to population DaR the demographic covariate values at time of relapse are considered.
 4. Remaining event times (time to relapse, excess NRM, and excess DaR) are simulated using $\lambda_i(t) = \lambda_0(t) \cdot \exp(\boldsymbol{\beta}^\top D_i)$.
 - In the simplest case, we take the hazard to be constant (i.e. we use the exponential distribution). In this case, the baseline hazard is constant and there are no covariate effects ($\lambda_0(t) = \lambda_0$).
 - In some simulating scenarios, we take Weibull distributed event times for these transitions. In this case, $\lambda_0(t)$ is the hazard of a Weibull distribution with parameters a (rate) and b (shape); $\lambda_0(t) = abt^{b-1}$. In this case, no covariate effects are considered. We further note that event times for the transition from relapse to excess DaR are simulated using a left-truncated Weibull distribution³³ where one conditions on the time of relapse; by this the Markov assumption is fulfilled.
 - Furthermore, we take simulation scenarios with a covariate effect (where $\boldsymbol{\beta}$ is not equal to zero). The covariate effect is included with respect to age, where age is first centered around 0.
 5. Although we simulate event times for both population and excess transitions, only the minimum of these is included in the dataset as observed value.
 6. The generated times are then subjected to censoring. Censoring times are simulated from the exponential distribution. The rate parameter of the distribution is modified for every scenario such that approximately 20% of all individuals are censored at the end of follow-up time (10 years).

Simulation scenarios: as mentioned, all population-related event times are generated using the same mortality tables, but population hazards vary between scenarios by adjusting the age distribution. Based on how the remaining event times are distributed and the magnitude of their hazards we define the five following simulation scenarios:

1. *exp.small*: for all transitions that are not population-related, the exponential distribution is used and the amount of population deaths is small.
2. *exp.large*: for all transitions that are not population-related, the exponential distribution is used and the amount of population deaths is close to the amount of excess deaths.
3. *weibull*: for all transitions that are not population-related, the Weibull distribution is used and the amount of population deaths is small.
4. *cov.eff.pos*: we allow for a covariate effect for certain transitions, the baseline hazard is again constant. The covariate effects of age on the hazard of having excess NRM and relapse are positive. No covariate effect is taken for the hazard of having excess DaR.
5. *cov.eff.neg*: we allow for a covariate effect for certain transitions, the baseline hazard is again constant. The covariate effect of age on the hazard of having excess NRM is positive, whereas on relapse it is negative. No covariate effect is taken for the hazard of having excess DaR.

Details on exact parameters can be found in the Supplemental Material. We additionally assess the impact of sample size on the estimators by varying sample size: $n = 500, 1000, \text{ and } 2000$ subjects.

3.1.3 Estimands

Our main measures of interest are the transition hazards and probabilities of the multi-state model. All values are assessed at 1, 2, 5, and 10 years, together with the corresponding SEs and CIs. All these measures have been defined in Section 2.3.

Exact transition hazards and probabilities (the true values) were calculated by integrating with respect to the covariate distribution D , which was not straightforward since times to population death were generated using mortality tables (see Section S1 in the Supplemental Material). The numerical integration was done in R.

3.1.4 Methods

We use the non-parametric approach outlined in Section 2.3 to estimate all transition hazards and probabilities, and we compare the performance of the variance estimators and the different CI methods.

3.1.5 Performance measures

Bias: let θ be the transition hazard or probability to be estimated at a given time for a given transition, whereas $\hat{\theta}_i$ is the estimate of θ for the i th replication of a simulation scenario. We define

$$\bar{\theta} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \hat{\theta}_i.$$

Then the definition of absolute and relative bias follows

$$\text{absolute bias} = \theta - \bar{\theta}, \quad \text{relative bias} = \frac{\theta - \bar{\theta}}{\theta}.$$

Relative bias is reported in the simulation results. This allows for an easy comparison of results across different timepoints.

Standard errors: we compared the two approaches for estimating variances (Greenwood and bootstrap) as defined in Section 2.3.2. For the bootstrap 100 replications are taken. We also calculate empirical SEs, that is, the standard deviation across all simulation replications

$$\widehat{\text{SE}} = \sqrt{\frac{1}{n_{sim} - 1} \sum_{i=1}^{n_{sim}} (\hat{\theta}_i - \bar{\theta})^2}.$$

Coverage probabilities: 95% CIs will be calculated for transition hazards and probabilities using the methods introduced in Section 2.3.3. The estimated CP is then the proportion of times the resulting CI contains the true estimand value.

We also use the CP for calculating the number of replications n_{sim} needed in the simulations. As in Section 5.3 in Morris et al.,³¹ we require that the SE for the CP equals 0.5%. Knowing that the expected CP equals 95%, it follows that $n_{sim} = 1900$. We note that if we would require the SE for the CP to be 1% we would obtain $n_{sim} = 475$. In our simulations, we round this to $n_{sim} = 2000$.

3.2 Results

3.2.1 Bias

Bias was negligible in all scenarios. Figure 4 shows the relative bias for cumulative transition hazards for all scenarios evaluated at the end of follow-up (10 years). An equivalent figure for transition probabilities is given in Figure S4 in the Supplemental Material. On average, relative bias was smaller when the sample size increased, which is expected. It seems that the relative bias was distributed around 0 and that behaviour was appropriate for all transitions and scenarios. For simulation scenarios with a covariate effect a slightly larger deviation from 0 was seen, but when the sample size was large enough (e.g. 2000), the relative bias was close to 0.

3.2.2 Standard errors

In Figure 5, we show SEs for scenario *exp.large* with sample size 2000. Through this example we will explain the crucial characteristics that were present in all of the simulation scenarios. The Greenwood option gave a zero estimate of the variance for population-related hazards and consequently, the estimated variance was also smaller for the corresponding transition probabilities. The bootstrap option better reflected the uncertainty in the estimates, as is confirmed by its values being close to the empirical SE. Furthermore, SEs were smaller for population-related transitions than for their excess counterparts. When the amount of population deaths was increased, the SE of these population-related transitions increased as well.

When looking at the left-truncated transition going from relapse to excess DaR, we can see that both variance estimators for the cumulative hazards, Greenwood and bootstrap-based, gave evidently smaller SEs than the empirical SE. This occurred for all simulation scenarios. As further explained in Section S3 of the Supplemental Material, this phenomenon does not only appear for the extended multi-state model but is already present in a basic multi-state model for left-truncated transitions. Solving this problem is out of the scope of this paper.

For transition probabilities (Figure 5) both estimator options were close to the empirical SE, the population-related transitions being an exception. In this case, the bootstrap was again a reliable option. The SEs for population-related transition probabilities increased when the amount of population deaths increased.

We find the bootstrap-based method a reliable option for estimating variances for all transitions in the wide range of scenarios we have explored. For some simulation scenarios (those with a covariate effect) there was a small difference

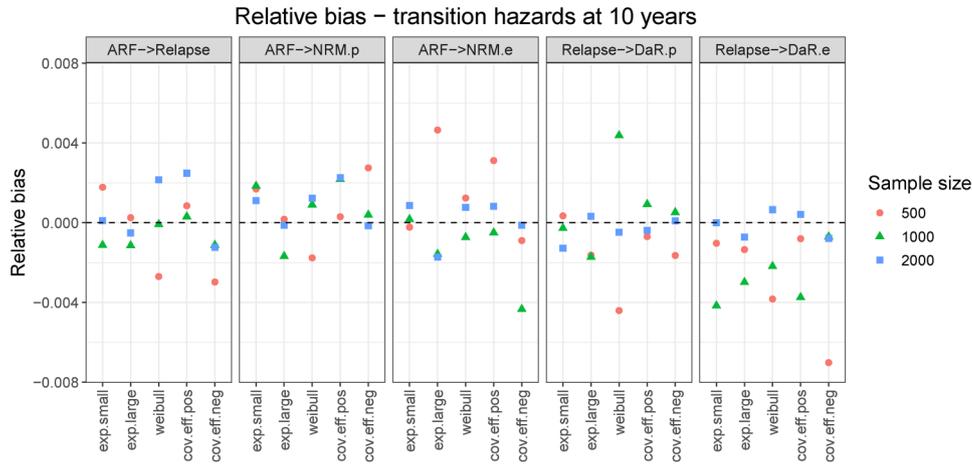


Figure 4. Relative bias for cumulative hazards evaluated at end of follow-up (10 years). States: Alive relapse-free (ARF), Relapse, population NRM (NRM.p), excess NRM (NRM.e), population DaR (DaR.p), and excess DaR (DaR.e).

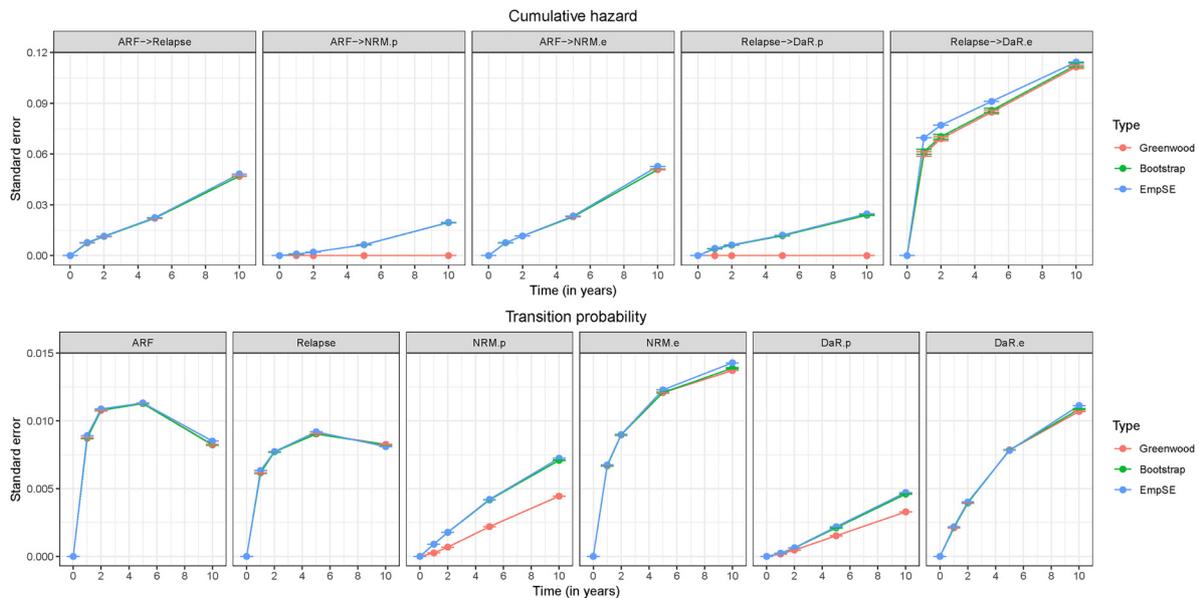


Figure 5. Standard errors for cumulative hazards and transition probabilities for scenario *exp.large* with sample size 2000 (average over 2000 replications). States: Alive relapse-free (ARF), Relapse, population NRM (NRM.p), excess NRM (NRM.e), population DaR (DaR.p), and excess DaR (DaR.e).

between the bootstrap estimate and the empirical SE for samples of size 500 and 1000. However, these differences diminished for samples of size 2000.

We additionally compared the SEs for the split transitions to the SEs for the observed transitions. In most cases, the empirical SEs for the observed transitions were very similar to the ones for the excess-related transitions. This has been proven to hold asymptotically for transition hazards, that is, $\text{Var}(\hat{\Lambda}_{hj}(t)) \approx \text{Var}(\hat{\Lambda}_{hj,E}(t))$.²² On the other hand, SEs for population-related transitions were smaller than for their excess-counterparts and they were mostly dependent upon the amount of population deaths and variability in the distribution of the demographic variables.

3.2.3 Confidence intervals

The SEs relative to the cumulative hazards and transition probabilities are small in either estimation approach. To additionally evaluate how well the two variance estimators (Greenwood and bootstrap) perform, we use CPs. Figure 6 shows CPs for transition probabilities at the end of follow-up. An equivalent graph for transition hazards is shown in the Supplemental Material (Figure S5). On average, CPs were more stable and closer to the nominal value for transition probabilities than for

hazards. By increasing the sample size the CPs also tended to increase and stay close to the nominal value (results not shown).

In almost all cases, *q.boot* gave smaller CPs than the other methods. For population-related transitions, the *plain.G* method gave inadequate CIs—on the hazard level we got zero-width CIs, on the probability level they were somewhat larger but their coverage was still far from the nominal value 95% (in many cases, this was smaller than 85% and thus not shown in Figure 6). This is a consequence of the variance estimator that was taken to equal 0 for these transitions. On the other hand, by using *plain.boot* (the same CI method but with the bootstrap SE estimate) we get much better coverages, see Figure 6. We also note that for transition probabilities, the *plain.G* and *plain.boot* methods can provide CIs that are outside of the [0, 1] interval.

The *plain.boot* and *log.boot* methods which were based on bootstrap SE estimates behaved similarly in the situations, but the latter method gave much better results for some cases. For left-truncated and population-related transitions the *log.boot* method gave larger CPs which were closer to the nominal coverage—this was especially evident at earlier time-points. An example of this is shown in Figure S6 in the Supplemental Material. Thus, *log.boot* is the preferred choice between the four introduced CI methods.

However, there were cases where even *log.boot* gave lower CPs than expected; this was most evident for some left-truncated and population-related transitions for scenarios *cov.eff.pos* and *cov.eff.neg*. This phenomenon occurred whenever the bootstrap SE estimate was lower than the empirical SE. To understand whether this is a consequence of the *log.boot* method we have also compared it to the *logit.boot* and *cloglog.boot* methods to see if an improvement in the CPs would occur. This has been considered in the Supplemental Material (Figure S9) where we can see that the three methods provide very similar CPs. Thus an improved variance estimator that does not underestimate the true variance (in these specific cases) is needed. We leave this question for further research.

4 Illustration

The new methodology is briefly illustrated by a reanalysis of previously presented data,⁴ which is taken as a motivating example in this article. The study population consisted of patients who had received a first allogeneic hematopoietic stem cell transplantation (alloHCT) for MDS or sAML between January 2000 and December 2012 and were recorded in the registry of the EBMT.³⁴ Further details on the selection criteria, patient characteristics and outcomes are available in the motivational paper.⁴

Patients were followed since alloHCT and possible outcomes were relapse/progression (for the remainder of this section we call this relapse) or death where we again distinguish between NRM and DaR. Thus, a multi-state model with a transition diagram as shown in Figure 1 is applicable, which can be further extended as shown in Figure 3. To better illustrate the usefulness of the methodology proposed, a subsample from the original dataset was taken, only including patients aged ≥ 60 years at alloHCT and still alive relapse-free at 2 years after alloHCT (the landmark time). Population mortality is

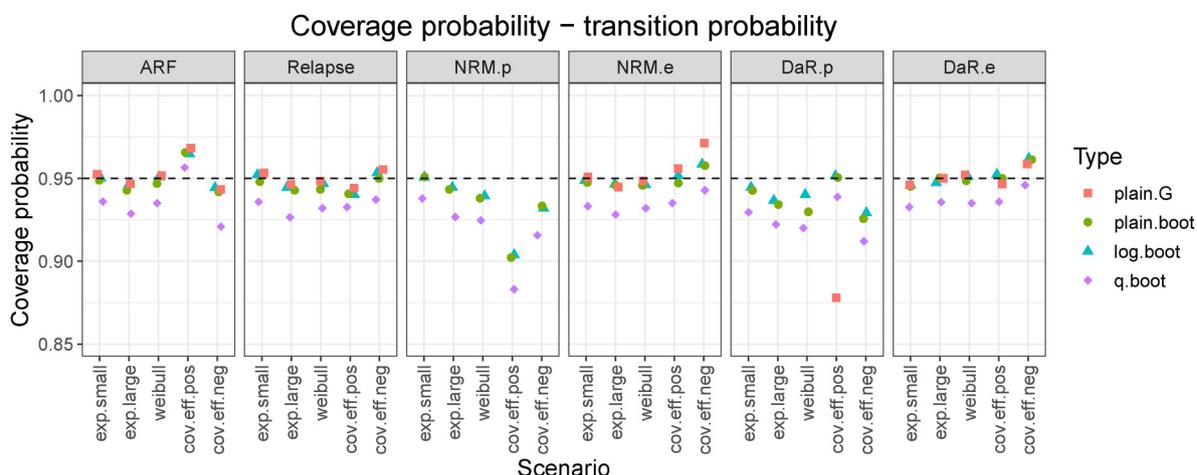


Figure 6. Coverage probabilities for transition probabilities evaluated at the end of follow-up (10 years) (averaged across sample sizes). Method *plain.G* gives inadequate coverage probabilities for population NRM and population DaR, many of which are smaller than 85% and thus are not shown in the figure. States: Alive relapse-free (ARF), Relapse, population NRM (NRM.p), excess NRM (NRM.e), population DaR (DaR.p), and excess DaR (DaR.e).

especially relevant for this group of older patients who have survived the first hazardous period after alloHCT. Data from a total of 753 patients from 19 European countries were included, with a median age of 63.9 years, where 62% of them were male and 38% female. For this example, we restrict follow-up time to 8 years.

Figure 7 shows the estimated cumulative hazards (left graph) and transition probabilities (right graph) for the extended multi-state model, whereas in the Supplemental Material, these estimated values are shown with the corresponding CIs, see Figures S7 and S8. The huge cumulative hazard of excess DaR is a sign of the very poor prognosis after relapse. The cumulative hazard of experiencing relapse is for most of the time slightly larger than the cumulative hazard of excess NRM. The hazards for the population-related transitions are very similar, their difference only caused by the different composition of the risk sets in terms of demographic variables in the states ARF and Relapse.

The transition probabilities (right graph in Figure 7) provide additional insight into the outcomes of the transplanted patients. The probability of being alive and relapse-free steadily decreases through time and reaches a probability of 56.9% (95% *log.boot* CI: [52.1, 62.1]) at 8 years (in this paragraph, all probabilities with the corresponding 95% *log.boot* CIs are reported at 8 years). The pattern shown by the hazards is propagated by the transition probabilities. The probability of dying due to population causes (population NRM or population DaR) equals 10.9%. The probability of population NRM equals 10.3% (95% CI: [9.7, 10.9]), whereas its excess counterpart (the probability of excess NRM) is 15.1% (95% CI: [11.3, 20.2]). On the other hand, the probability of population DaR is much smaller than of excess DaR: 0.6 (95% CI: [0.4, 1.0]) and 13.3 (95% CI: [10.7, 16.5]), respectively, which is due to the strong competing risk of excess DaR. The probability of being in the relapse state (and having no further event) is steady after 2 years reaching 3.9% (95% CI: [1.7, 8.8]) at end of follow-up.

The clinical interpretation of these results is that DaR is almost identical to death due to relapse. In the past, only young patients were referred for alloHCT. For these younger patients (below the age of 60 years at the time of alloHCT) NRM is almost exclusively treatment-related mortality (where treatment refers both to the alloHCT and to other treatments before and after). Yet age at HCT has continuously increased in the past decades, and the presumption that NRM equals treatment-related mortality is not valid for the older transplanted population. In our example, this is demonstrated by the fact that 40.6% of NRM at 10 years after alloHCT for this landmark cohort can be attributed to population mortality. Thus, their treatment-related mortality can better be estimated by excess NRM.

The R package `mstate`³⁰ v0.3.2 allows for the introduced extension. After the transition hazards are estimated using the function `msfit`, they are supplied to the newly developed `msfit.relsurv` function. This produces a new and extended `msfit` object containing the hazards for the excess and population transitions and their variances, calculated by the methods described in Section 2.3.2. The extended `msfit` object can then be supplied to function `probtrans`, through which transition probability estimates and their variance estimates are obtained. Furthermore, plain and log CIs (as defined in Section 2.3.3) can be obtained using the `summary.msfit` and `summary.probtrans` functions. Thus the original package has been upgraded

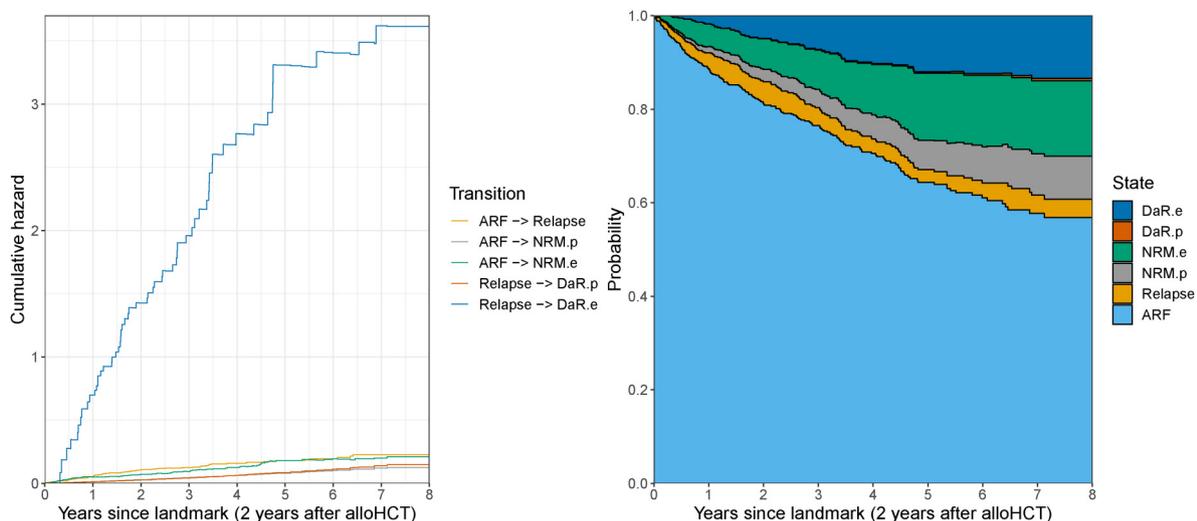


Figure 7. Estimated cumulative hazards (left) and transition probabilities (right) for patients 65 years or older at alloHCT and alive relapse-free at the 2-year landmark time estimated in the EBMT dataset (as introduced in Section 4). Estimates provided up to 8 years since the landmark time. States: Alive relapse-free (ARF), Relapse, population NRM (NRM.p), excess NRM (NRM.e), population DaR (DaR.p), and excess DaR (DaR.e).

in such a way that only a small additional step is needed when coding (function `msfit.relsurv`). A full code example using `mstate` is provided in Section S4 of the Supplemental Material.

5 Discussion

In this paper, we have introduced an extension of a multi-state model containing death states where the exact cause of death is not known. We were able to account for death by other (population) causes in such a model using external population mortality tables. The relative survival theory has allowed for such an extension where we have solely focused on non-parametric estimation in a Markov setting. Thus, we have successfully integrated the basic relative survival idea (of splitting a death state in excess and population counterparts) in a general multi-state framework for any transition reaching death.

The present work focuses on transition hazards (through which the extension has been done) and transition probabilities, the latter being more easily interpretable for a non-statistical audience. The proposed estimators are defined in such a way that the information from mortality tables is properly taken into account. Based on our simulation study we can conclude that as expected, estimation bias is very minor.

By using the extended multi-state model we also have to assume the additive model (equation (6)) for all split transitions. Thus, in order for the excess hazard to be a proper hazard, it needs to be positive across the whole time interval. There are many practical settings where this does not hold³⁵ because the population under study has a better life expectancy than the general population, for example, because primarily patients with a more robust general health might survive certain diagnoses. If the model is still applied even though it does not really hold, *negative* cumulative hazards and transition probabilities are obtained, which could still give a useful message, if interpreted with caution. In such a case, negative cumulative hazards can only be obtained for excess-related transitions which then affect the transition probabilities of the excess and population death states.

Throughout the paper, we only split transitions to death states. However, the methods (together with the software implementation) allow transitions reaching intermediate events to be split as well. In this case, we would need adequate population tables for those specific transitions. For example, if we want to examine if patients with an autoimmune disease run a larger risk of a COVID-19 infection and death thereafter than the general population, a model could be developed with starting state diagnosis, intermediate event COVID-19 infection and absorbing states death before and after COVID-19 infection. Based on observed data for a cohort of these patients and population tables for the hazard of infection by COVID-19 and death after infection, the transition to the intermediate event could be split into a population and an excess part and similarly the transition from infection to death after infection could be split. A different example is available¹⁵ where a model for Hodgkin lymphoma patients with a disease of the circulatory system (DCS) as an intermediate event is considered.

The present work has concentrated on non-parametric estimation of transition hazards, from which the relevant transition probabilities and state occupation probabilities can be derived. Covariates like age and sex have only been used for providing the correct population hazards. A next natural step would be to incorporate covariates in the estimation of the (excess) transition hazards. The most common approach is the Cox proportional hazards model, which has been used in multi-state models²⁴ and in relative survival.³⁶ Modeling by additive hazards is also an attractive alternative,³⁷ since the distinction between population and excess hazards for a particular transition is also additive, and the subsequent machinery of calculating transition probabilities from the transition hazards also involves addition of the transition hazards. Other modeling approaches have been also considered in previous literature.^{11–15}

The suggested theory only uses a small part of the field of relative survival; it does not consider any other relative survival measures, for example, net survival or the relative survival ratio, and their extension to a general multi-state model. We leave this point for further research. We note that the crude probability of death (a common measure in relative survival) has its analogon in the split transition probabilities $\hat{P}_{hj,P}(s, t)$ and $\hat{P}_{hj,E}(s, t)$.

As in any multi-state model, one has to carefully check the Markov assumption when modeling.³⁸ For a general multi-state model, there can be intermediate states for which the Markov assumption does not hold. One such example would be if the transition hazard of reaching DaR from relapse is dependent on the duration time in the relapse state. Estimation for Markov renewal models where transition hazards depend on the duration time has been previously discussed.³⁹ Since for Markov renewal models the relevant time scale is time since entry into the current state the population hazard in equation (9) can still be used, but times t and u should be understood to represent time since entry in state h . If the multi-state model is neither Markov nor Markov renewal, in the setting without population mortality, the estimators of the transition hazards (in the non-Markov setting) are consistent, as a consequence of Datta and Satten.²⁰ The same authors also show that the state occupation probability estimator derived from equation (5) is consistent from which it has been proved that the landmark Aalen-Johansen estimator²¹ is also consistent. Unfortunately, we think that these results do not extend to the relative

survival case in a straightforward way, since the population hazards are not estimated internally together with the other transition hazards but are coming from external sources, and hence the result of Datta and Satten²⁰ does not apply here. On the other hand, it should be noted that simulation studies have shown that often estimates of transition probabilities are remarkably robust to violations of the Markov assumption.²¹ Hybrid approaches like suggested in Maltzahn et al.⁴⁰ could well be reasonable compromises in the sense of sacrificing a small amount of bias in favour of reduced variance. This needs to be investigated in future studies. Another approach for estimating transition probabilities would be using plug-in models.⁴¹ However, this approach requires further assumptions which we do not consider in this work.

For estimating the variance of the proposed measures, we suggest the bootstrap approach which was a suitable option in most of the simulation scenarios. The Greenwood estimator was also considered as a closed-form estimator. This required the additional assumption for the variance estimates of the population-related transition hazards to be equal to 0. This assumption proved to be inadequate since the population-related transitions have substantial variability which is not considered by this approach. For the remaining transitions the Greenwood estimator proved to be a suitable option. Instead of making this additional assumption (using the Greenwood estimator), a non-parametric variance estimator for population-related transitions would be of great use. However, such an estimator has not been developed (as far as we know). Developing such a variance estimator for the population-related transitions could find its use for many measures related to relative survival and we leave this for future research.

An additional problem with variance estimation of left-truncated transitions has been also pointed out (and further illustrated in the Supplemental Material). This problem is inherent to a basic multi-state model already and is out of the scope of this paper but can be tackled in future work.

In the simulation study, we have also considered CI methods. The *log.boot* proved to be a reliable choice across simulations, and is therefore our recommended option. There were cases when the CP of *log.boot* was evidently smaller than the nominal value. For a comparison, we have also checked whether the additional *logit.boot* and *cloglog.boot* methods would provide better CPs for the cases when the CP was below 0.95. The three methods have provided very similar results. Thus a further improvement in the CP would require developing a new variance estimator which does not underestimate the true variance.

All work has been implemented in R and is free for use through the *mstate* package.³⁰ The implementation is based on functions from package *relsurv*²⁸ which has allowed for an easy and optimal usage of mortality tables which is otherwise not trivial. The current R implementation also allows for non-parametric estimation based on left-truncated observations in a standard survival setting (which has not been always considered in previous software implementations) and for future extensions to Cox-model based semi-parametric models. From what we know this is the first such implementation in R. Its use has been illustrated through an application. The Supplemental Material shows self-contained code that can help the reader to apply this methodology.

To conclude, we believe that the proposed extension of a non-parametric multi-state model could find its use in a wide range of practical applications. Especially for interventions encompassing potential long-term beneficial and adverse effects offered to older patients with a remaining life-expectancy of several years or even a decade, taking into account population mortality as part of the different components of mortality is relevant. In this paper, we have mostly focused on the theoretical aspects of this work. By conducting a thorough simulation study, we have further understood the properties of the estimators and come across additional challenges, some of which have to be dealt with in the future (e.g. variance estimation for left-truncated transitions, a theoretical variance estimator for population-related transitions, an improved CI method). We have devised an R application that relies on previous multi-state implementations in R. Thus by providing such a tool we believe that this work can be beneficial in many practical settings.

Acknowledgements

We thank the EBMT Data Office in Leiden for providing the MDS dataset.

Authors' note

Johannes Schetelig and Liesbeth C de Wreede, Clinical Trials Unit, DKMS, Dresden, Germany.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

Damjan Manevski is a young researcher funded by the Slovene Research Agency (ARRS). The authors acknowledge the financial support by ARRS (Methodology for data analysis in medical sciences, P3-0154 and project J3-1761).

ORCID iDs

Damjan Manevski  <https://orcid.org/0000-0001-9044-2862>

Hein Putter  <https://orcid.org/0000-0001-5395-1422>

Edouard F Bonneville  <https://orcid.org/0000-0001-7542-4498>

Supplemental material

Supplemental material for this article is available online.

References

- Mariotto AB, Noone AM, Howlander N et al. Cancer survival: An overview of measures, uses, and interpretation. *J Natl Cancer Inst Monogr* 2014; **2014**: 145–186.
- Begg CB and Schrag D. Attribution of deaths following cancer treatment. *J Natl Cancer Inst* 2002; **94**: 1044–1045.
- Percy C, Stanek E and Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. *Am J Public Health* 1981; **71**: 242–250.
- Schetelig J, de Wreede LC, van Gelder M et al. Late treatment-related mortality versus competing causes of death after allogeneic transplantation for myelodysplastic syndromes and secondary acute myeloid leukemia. *Leukemia* 2019; **33**: 686–695 .
- Pohar Perme M, Stare J and Estève J. On estimation in relative survival. *Biometrics* 2012; **68**: 113–120.
- Human mortality database, 2016. Available at www.mortality.org www.humanmortality.de.
- Ederer F, Axtell LM and Cutler SJ. *The relative survival rate: a statistical methodology*, volume 6. National Cancer Institute Monograph, 1961. pp. 101–121.
- Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 1982; **38**: 933–942.
- Sasieni P and Brentnall AR. On standardized relative survival. *Biometrics* 2016; **73**: 473–482 .
- Pavlič K and Pohar Perme M. Using pseudo-observations for estimation in relative survival. *Biostatistics* 2018; **20**: 384–399.
- Husztí E, Abrahamowicz M, Alioum A et al. Relative survival multistate Markov model. *Stat Med* 2012; **31**: 269–286.
- Belot A, Remontet L, Launoy G et al. Competing risk models to estimate the excess mortality and the first recurrent-event hazards. *BMC Med Res Methodol* 2011; **11**: 7–8.
- Gillaizeau F, Dantan E, Giral M et al. A multistate additive relative survival semi-Markov model. *Stat Methods Med Res* 2017; **26**: 1700–1711.
- Crowther MJ and Lambert PC. Parametric multistate survival models: Flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences. *Stat Med* 2017; **36**: 4719–4742.
- Weibull CE, Lambert PC, Eloranta S et al. A multistate model incorporating estimation of excess hazards and multiple time scales. *Stat Med* 2021; **40**: 2139–2154.
- 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 .
- Allignol A, Schumacher M and Beyersmann J. A note on variance estimation of the AalenJohansen estimator of the cumulative incidence function in competing risks, with a view towards left-truncated data. *Biom J* 2010; **52**: 126–137.
- Friedrich S, Beyersmann J, Winterfeld U et al. Nonparametric estimation of pregnancy outcome probabilities. *Ann Appl Stat* 2017; **11**: 840–867.
- Rousson V, Allignol A, Arousseau A et al. Stabilizing cumulative incidence estimation of pregnancy outcome with delayed entries. *Biom J* 2019; **61**: 1290–1302.
- 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.
- Putter H and Spitoni C. Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark AalenJohansen estimator. *Stat Methods Med Res* 2018; **27**: 2081–2092.
- Andersen PK, Borgan O, Gill RD et al. *Statistical models based on counting processes*. New York: Springer-Verlag, 1993.
- Andersen PK and Keiding N. Multi-state models for event history analysis. *Stat Methods Med Res* 2002; **11**: 91–115.
- Putter H, Fiocco M and Geskus RB. Tutorial in biostatistics: Competing risks and multi-state models. *Stat Med* 2007; **26**: 2389–2430.
- Klein JP, van Houwelingen HC, Ibrahim JG et al. (eds.), *Handbook of survival analysis*. Chapman & Hall/CRC handbooks of modern statistical methods, Boca Raton, FL: CRC Press, 2014. ISBN 9781466555662.
- Tsiatis AA. *Competing Risks*. Chichester, UK: Wiley, 2005. ISBN 9780470011812.
- Andersen PK and Keiding N. Interpretability and importance of functionals in competing risks and multistate models. *Stat Med* 2012; **31**: 1074–1088.
- Pohar Perme M and Pavlič K. Nonparametric relative survival analysis with the R package relsurv. *J Stat Soft, Articles* 2018; **87**: 1–27.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. <https://www.R-project.org/>.
- de Wreede L, Fiocco M and Putter H. mstate: An R package for the analysis of competing risks and multi-state models. *J Stat Soft, Articles* 2011; **38**: 1–30.
- Morris TP, White IR and Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med* 2019; **38**: 2074–2102.

32. Meira-Machado L and Faria S. A simulation study comparing modeling approaches in an illness-death multi-state model. *Commun Stat – Simulat Comput* 2014; **43**: 929–946.
33. Wingo DR. The left-truncated Weibull distribution: Theory and computation. *Stat Papers* 1989; **30**: 39–48.
34. The EBMT Patient Registry. <https://www.ebmt.org/ebmt-patient-registry>.
35. Antero-Jacquemin J, Pohar Perme M, Rey G et al. The heart of the matter: Years-saved from cardiovascular and cancer deaths in an elite athlete cohort with over a century of follow-up. *Eur J Epidemiol* 2018; **33**: 531–543.
36. Pohar Perme M and Stare J. Relative survival analysis in R. *Comput Methods Programs Biomed* 2006; **81**: 272–278.
37. Zahl PH. A linear non-parametric regression model for the excess intensity. *Scand J Stat* 1996; **23**: 353–364.
38. Titman AC and Putter H. General tests of the Markov property in multi-state models. *Biostatistics* 2020; **10**: 030.
39. 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.
40. Maltzahn N, Hoff R, Aalen OO, et al. A hybrid landmark aalen-johansen estimator for transition probabilities in partially non-Markov multi-state models. *Lifetime Data Anal* 2021; **27**: 737–760.
41. Andersen PK and Pohar Perme M. Inference for outcome probabilities in multi-state models. *Lifetime Data Anal* 2008; **14**: 405.