Journal Title
XX(X):1–31

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

A new framework for semi-Markovian parametric multi-state models with interval censoring

Marthe Elisabeth Aastveit[1*], Céline Cunen[1,2*] and Nils Lid Hjort[2]

Abstract

There are few computational and methodological tools available for the analysis of general multi-state models with interval censoring. Here, we propose a general framework for parametric inference with interval censored multi-state data. Our framework can accommodate any parametric model for the transition times, and covariates may be included in various ways. We present a general method for constructing the likelihood, which we have implemented in a ready-to-use R package, `smms`, available on GitHub. The R package also computes the required high-dimensional integrals in an efficient manner. Further, we explore connections between our modelling framework and existing approaches: our models fall under the class of semi-Markovian multi-state models, but with a different, and sparser parameterisation than what is often seen. We illustrate our framework through a dataset monitoring heart transplant patients. Finally, we investigate the effect of some forms of misspecification of the model assumptions through simulations.

Keywords

competing risk, interval censoring, multi-state models, panel data, semi-Markov models, survival analysis, time-to-event

1. Introduction

Multi-state models are widely used in biostatistical applications to model phenomena where the units of observation, typically patients, transition through a set of discrete states on their way towards one or

Corresponding authors:
Marthe Elisabeth Aastveit, Norwegian Computing Center, Gaustadalléen 23A, 0373 Oslo, Norway. Email: aastveit@nr.no
Céline Cunen, Norwegian Computing Center, Gaustadalléen 23A, 0373 Oslo, Norway. Email: celine@nr.no

Prepared using sagej.cls *[Version: 2017/01/17 v1.20]*

more absorbing states. The states may constitute various stages of a disease, from perfect health through various stages of dementia for example. In many applications the transition times between states are not observed exactly; instead, the current state of the patients is queried at arbitrary times. These types of data are often referred to as panel data [1]. The transition times are then interval censored, and this makes inference and modelling challenging. There are few methodological and computational tools available for a user studying a multi-state phenomenon where the observation times are interval censored. If the user is willing to assume that the phenomenon is Markovian, i.e. that the future depends only on the present state, there are computational tools available, primarily the `msm` package in R [2]. Further, if the transition times are not interval censored, but observed exactly, there exist several frameworks accommodating models which relax the Markov assumption [3–5]. In addition, some papers have worked out the likelihood for semi-Markov multi-state models with interval censored data [6–10], but most of those papers treat particular cases or datasets, and none provide general computational implementations of their methods. This is a limitation which we have sought to address. In this paper we propose a novel modelling framework for multi-state data with interval censoring, we describe a general algorithm for constructing and maximising the log-likelihood, and we explore connections with other semi-Markovian models in the literature.

Our framework can accommodate a wide range of different parametric models for the transition times. Further, covariates may influence the transition times in various and flexible ways. The user needs to provide two sets of inputs: (1) the multi-state *structure* in the form of a graph, which provides the possible transitions between states, see for example Figure 1; (2) a set of parametric models for the transition times, which can be of various types, as we will come back to. In this paper, we limit ourselves to multi-state models without cycles, i.e. directed acyclic graphs (DAGs). In the multi-state literature such graphs may be referred to as unidirectional models [7]. Even with this limitation we still cover a wide range of different real-life phenomena, for example irreversible diseases like dementia [11], HIV [12] and cirrhosis [13]. Our framework can be extended to accommodate cycles, but this requires additional assumptions - we will briefly come back to this potential extension in the discussion. Apart from this limitation, we do not have further restrictions on the types of graph we can accommodate. We assume that the graph has one, or more, initial states, and one or more absorbing states. We allow for unobserved transitions, meaning that some patients may have transitioned through several states without being observed in them. For some graphs, this introduces the possibility of multiple potential paths, which increases the complexity of the likelihood, as we will see. We assume that the observation process is independent of the multi-state process, and that there exist a natural and common time-origin for all patients, see discussions by Commenges [14] and by Aralis and Brookmeyer [10].

Contrary to most of the literature, we prefer to construct our models based on the transition densities, rather than the transition intensities as is commonly done. Nonetheless, we obtain models that fall into a known class, i.e. that of semi-Markov models. Semi-Markov models relax the Markov assumption by allowing future transition probabilities to depend on the time spent in the current state [15]. There exists two general types of semi-Markov models in the literature, and our approach belongs to the one leading to a smaller number of parameters. We will explore the relationship between the two types of semi-Markov models in Section 4 and also in the simulations.

The rest of the paper is structured as follows. In Section 2 we describe our modelling framework: the construction of the likelihood in a simple example first, and then the general recipe for tackling a general multi-state graph. In Section 3 we briefly describe our algorithm and computational choices. Note here that the interval censoring leads to high-dimensional integrals which constitute a major computational

challenge. Further, in Section 4 we draw connection between our approach and the simple homogeneous Markov model, and between our approach and other semi-Markov models. We then illustrate the use of our framework through an application in Section 5. In Section 6, we present results from simulations where we have studied the effect of some forms of model misspecification: first, to what degree wrongly assuming a simple Markov model leads to erroneous statements, and next whether the choice between the two semi-Markov approaches changes our inference to a large degree. Finally, we discuss limitations and potential extensions of our framework in Section 7.

2. Methodology

2.1. The Framework

To each edge in a multi-state graph we assign a continuous random variable, T_{ij}, with i indicating the current state and j indicating the potential next state. See for example Figure 1 representing the three-state illness-death model. There we have three random variables T_{01}, T_{02} and T_{12}. In the following, we will refer to these random variables as *transition times*, but a more precise definition is the time from entering state i to the transition to state j occurs, or would have occurred. The definition indicates that we let the random transition times compete with each other to determine the next state. Say for example we have the illness-death model from Figure 1 and a patient who is currently in state 0. If $T_{01} < T_{02}$ the patient will proceed to state 1, while in the opposite event the patient will proceed to state 2. The time from the time origin to entering the absorbing state will be the sum of transition times belonging to the path the patient has travelled, for instance $T_{01} + T_{12}$ for a patient having reached the absorbing state via state 1. All patients are assumed to enter the initial state at time zero.

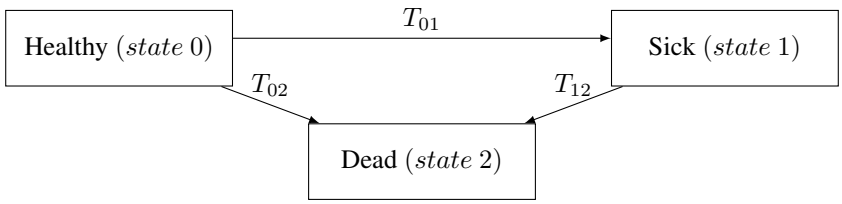


Figure 1. A three-state illness-death model

The transition times T_{ij} are latent and may refer to transitions that do not in fact take place (if the patient takes a different path). They are assumed to be independent and drawn from distributions over the positive half-line,

$$T_{ij} \sim f_{ij}(\cdot\,; \theta_{ij})$$

with density functions $f_{ij}(t; \theta_{ij})$ and transition-specific parameter vector θ_{ij}. For ease of notation, we will often omit θ_{ij} in the following. The different $f_{ij}(t; \theta_{ij})$ may come from different classes of parametric distributions, like Weibull, Gamma or even threshold-hitting-models like the Gamma processes threshold-models used by Aastveit[16]. We will denote the associated survival functions by $S_{ij}(t)$, and the transition intensity functions by $h_{ij}(t)$. These may be interpreted in similar ways as

hazard functions, but with the caveat that they relate to events that might not take place. As usual, we have $h_{ij}(t) = f_{ij}(t)/S_{ij}(t)$.

We will index the units of observations, which we will refer to as patients in the rest of the paper, by $k = 1, \ldots, n$. If covariates x_k have been recorded, these variables may be allowed to influence the transition times. This can be achieved in various ways. One option is to let some or all model parameters θ_{ij} be functions of the covariates, for example by

$$\theta_{ij,k} = \exp(x_k^{\mathrm{t}}\beta_{ij}), \quad k = 1, \ldots, n,$$

with x_k the vector of covariates for patient k and β_{ij} the corresponding vector of coefficients (here we assume that θ_{ij} is a one-dimensional parameter; if it is multi-dimensional the user must choose whether to let covariates influence all parameters, or only some). Other link functions than the log link may be considered depending on the situation. Another option is to include covariates on the scale of the transition intensity functions, assuming proportionality:

$$h_{ij,k}(t) = h_{ij,0}(t)\exp(x_k^{\mathrm{t}}\beta_{ij}), \quad k = 1, \ldots, n.$$

The user may choose to let covariates influence all, or only a selected subset of transitions. This choice may be guided by model selection criteria.

For each patient one has recorded a number of observations of state occupancy, at *observation times* t_r, with $r = 1, \ldots, m_k$. The observation times live on the common time-scale, starting at zero when the patient entered the initial state(s), while the transition times have their own time-scales, starting at zero upon entering a new state. Now we will derive the likelihood in a simple example, before giving the general recipe.

2.2. The likelihood for the three-state illness-death model

A frequently studied multi-state structure is the illness-death model illustrated in Figure 1. Here we will present the likelihood for that model in a situation where we assume that all transitions are interval censored. For the graph in Figure 1, four combinations of states may be observed, '0', '01', '02' and '012'. Each patient belongs to one of these four combinations: '0' denotes patients who are only observed in state 0, '01' patients who are observed in state 0 and then in state 1, and so on. The log-likelihood will then be the sum of four different types of likelihood contributions,

$$\ell_n(\theta) = \sum_{k=1}^{n_0} \log L_{0,k}(\theta) + \sum_{k=1}^{n_{01}} \log L_{01,k}(\theta) + \sum_{k=1}^{n_{02}} \log L_{02,k}(\theta) + \sum_{k=1}^{n_{012}} \log L_{012,k}(\theta),$$

with n_0 indicating the number of patients observed only in state 0, and so on; θ indicates the full parameter vector for the model under consideration. It follows directly from the framework described in the previous section that the likelihood contribution of a patient only observed in state 0 can be given as,

$$L_{0,k}(\theta) = \Pr(T_{01} > t_{0M}, T_{02} > t_{0M}) = S_{01}(t_{0M})S_{02}(t_{0M}),$$

where t_{0M} is the last observed time-point in state 0 (these time-point are typically different for each patient, and could therefore have been denoted as $t_{0M,k}$, but we let these indices be implicit). The next

type represents patients which are observed in state 0 from time-point 0 to t_{0M}, then in state 1 at a later time-point t_{1m},

$$\begin{aligned} L_{01,k}(\theta) &= \Pr(t_{0M} < T_{01} < t_{1m}, T_{01} + T_{12} > t_{1M}, T_{02} > T_{01}) \\ &= \int_{t_{0M}}^{t_{1m}} f_{01}(s) S_{12}(t_{1M} - s) S_{02}(s)\, ds, \end{aligned} \tag{1}$$

where t_{1M} is the last observed time-point in state 1. Here, we have to integrate over the unobserved transition time T_{01}. The third type represents patients which are observed in state 0 from time-point 0 to t_{0M}, and in state 2 at a later time-point t_{2m}. In this case, the combination of observed states must have originated from one out of two possible paths in the graph. The first possibility is that the individual transitioned directly from state 0 to state 2, the second that the individual transitioned through state 1, but was not observed there. Because of these two possibilities the likelihood contribution consists of the sum of two probabilities.

$$\begin{aligned} L_{02,k}(\theta) &= \Pr(t_{0M} < T_{01} < t_{2m}, t_{0M} < T_{01} + T_{12} < t_{2m}, T_{02} > T_{01}) \\ &\quad + \Pr(t_{0M} < T_{02} < t_{2m}, T_{01} > T_{02}) \\ &= \int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} f_{01}(s) f_{12}(u) S_{02}(s)\, du\, ds + \int_{t_{0M}}^{t_{2m}} f_{02}(s) S_{01}(s)\, ds \\ &= \int_{t_{0M}}^{t_{2m}} f_{01}(s)[1 - S_{12}(t_{2m} - s)] S_{02}(s)\, ds + \int_{t_{0M}}^{t_{2m}} f_{02}(s) S_{01}(s)\, ds \end{aligned} \tag{2}$$

The inner integral in the first probability can be written out explicitly as $[S_{12}(0) - S_{12}(t_{2m} - s)]$. Finally, we have patients which are observed in all three states: in state 0 from time-point 0 to t_{0M}, then in state 1 from time t_{1m} to t_{1M}, and finally in state 2 at time t_{2m},

$$\begin{aligned} L_{012,k}(\theta) &= \Pr(t_{0M} < T_{01} < t_{1m}, t_{1M} < T_{01} + T_{12} < t_{2m}, T_{02} > T_{01}) \\ &= \int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} f_{01}(s) f_{12}(u) S_{02}(s)\, du\, ds \\ &= \int_{t_{0M}}^{t_{1m}} f_{01}(s)[S_{12}(t_{1M} - s) - S_{12}(t_{2m} - s)] S_{02}(s)\, ds \end{aligned} \tag{3}$$

Again, the inner integral can be written out explicitly in this case.

In many applications, the transition time into the absorbing state is not censored, but observed exactly. This is typically the case when the absorbing state represents death of the patient, whose timing is often recorded precisely. In the example here this entails that realisations of T_{12} and T_{02} are observed exactly and that we no longer need to integrate over these random variables. Specifically, we would need to replace $1 - S_{12}(t_{2m} - s)$ and $S_{12}(t_{1M} - s) - S_{12}(t_{2m} - s)$ in Equations (2) and (3) by $f_{12}(t_2 - s)$, where t_2 is the exact time of entry into state 2. Further, the second integral in Equation (2) would simply be replaced by its integrand evaluated in t_2.

In the appendix, we provide the likelihood for the four-state illness-death, and for some other multi-state graphs. These are meant to illustrate the large variety in these expressions and the need for a general recipe. The four-state illness-death model for example has eight types of likelihood contributions, each being the sum of up to three terms.

2.3. The general recipe

As the illustration in Section 2.2 indicates, the likelihood is built up in a systematic manner. The integrands are products of densities and survival functions, and the integral limits come from the observation time-points bounding the unobserved transition times. We can therefore derive a set of rules for the construction of the likelihood belonging to a specific multi-state graph.

1. Enumerate all possible combinations of states that one might observe. These define the likelihood contribution types, for example '0', '01', '02' and '012' as we saw in Section 2.2.
2. Determine which of these combinations contain multiple possible paths. If there is more than one possible path, the contribution will be a sum of terms corresponding to each path. Each term will be an integral (possibly of dimension zero in which case it is simply a function to evaluate). The dimension of the integral is equal to the number of edges that have been travelled, so for example type '0' leads to a zero dimensional integral, while type '012' leads to a two dimensional integral (with the possibility of some simplification as we will see in step 5.).
3. Construct the integrand for each given path. It will be the product of:
   (a) The density functions of the variables belonging to edges that have been *travelled*. These densities are evaluated at the (unobserved) transition time belonging to the travelled edge. See for example the term $f_{02}(s)$ in Equation (2), where s stands for the unobserved T_{02}, which is integrated over.
   (b) The survival functions of the variables belonging to edges that could have been travelled, but were instead *passed-by*. These survival functions are evaluated at the (unobserved) transition time belonging to the edge that was travelled instead of the one which was passed-by. See for example the term $S_{01}(s)$ in Equation (2), where s again stands for the unobserved T_{02}, which is integrated over.
   (c) The survival functions of the variables belonging to edges that *may be travelled in the next step*. If the edge between state i and j may be travelled in the next step, the corresponding survival functions will be evaluated at t_{iM}, the last observation time-point in state i, minus all the variables belonging to edges that have been travelled before state i. See for example the term $S_{12}(t_{1M} - s)$ in Equation (1).
4. Determine the limits of integration based on the time-points bounding each unobserved transition time, for example in Equation (3) T_{01} is bounded by t_{0M}, the last observed time-point in state 0, and t_{1m}, the first observation in state 1. Further, T_{12} is bounded by $t_{1M} - s$ and $t_{2m} - s$, where s stands for the unobserved T_{01}, which is integrated over in the outer integral. Often, the lower limit will correspond to negative values, since two subsequent transitions are bounded by the same limits, see for example Equation (2). In that case we replace the lower limit by zero.
5. Sometimes, the innermost integral can be written out explicitly, as we saw in Equations (2) and (3). This happens when an absorbing state is reached from a state where there is only a single option, meaning that the next-to-last state only has a single edge going out of it. This is a quite common situation, but see Appendix B for a situation where this type of simplification is not possible.

If some of the transition times, typically the one belonging to the entrance into the absorbing state, are observed exactly, the likelihood must be modified in a similar manner to what is explained in the end of Section 2.2. We show some examples of such likelihoods in the appendix. Further, if all

transition times are observed exactly, then the likelihood simplifies to a well-known form, found in several contributions [10,15,17], consisting of the product of density and survival functions, without any integrals.

2.4. Maximum likelihood estimation

Once we have workable expressions for the log-likelihood functions, these may be maximised with respect to the parameters, yielding maximum likelihood (ML) estimators $\hat\theta$ for the full parameter vector θ of a given multi-state model. The traditional theory works for these, under reasonable regularity conditions, via suitable extensions of standard theory to the present cases with interval censored data. For the special case of there being a fixed finite set of intervals, where one for each transition time T in question knows in which of these intervals T falls, the setup can be translated to a multinomial setting, involving cell probabilities, say $p_{i,j}(x_i,\theta)$ for individuals i and cells j. This again leads to a clear log-likelihood function where standard maximum likelihood theory applies; see e.g. chapters 2 and 3 in Claeskens and Hjort [20]. More general setups require more elaborate extensions; see e.g. [18,19].

Specifically, assuming the model holds, with θ_0 the true underlying parameter value, of dimension p, the ML estimator $\hat\theta$ is approximately multinormal $\mathrm{N}_p(\theta_0,\hat J^{-1})$, where $\hat J=-\partial^2\ell_n(\hat\theta)/\partial\theta\partial\theta^{\mathrm{t}}$ is the observed Fisher information matrix. This leads upon using the delta method also to approximate normality of any consequent estimator $\hat\mu=\mu(\hat\theta)$, for any focus parameter quantity $\mu=\mu(\theta)$, along with a consistent estimate of its standard deviation. Furthermore, model selection methods and theory associated with the AIC and the BIC (the Akaike and Bayesian information criteria) will also work essentially as usual, see Claeskens and Hjort [20]. In our `smms` package we fit models by maximum likelihood, and we use the theory alluded to here for inference and model selection, see also Section 3. The approximate normality of the estimators in this type of multi-state model was investigated extensively in Aastveit [16], for several different graphs. We also include a brief simulation in Section 6.3 where we check whether the ordinary approximate confidence intervals obtain the correct coverage rates.

2.5. Quantities of interest

After having fitted a multi-state model, one may want to compute various quantities of interest both for the sake of interpreting the fitted model, and also for discovering potential lack of fit. Here we will briefly present three such quantities, (i) the overall survival curve, (ii) the state occupancy probabilities, and (iii) the transition probabilities.

The overall survival curve gives the probability of not having reached an absorbing state at time t. For the simple three-state illness-death model from Figure 1, we get

$$\begin{aligned}\Pr(\text{survival}) &= \Pr(T_{01}>t,T_{02}>t)+\Pr(T_{01}+T_{12}>t,T_{01}<t,T_{01}<T_{02}) \\ &= S_{01}(t)S_{02}(t)+\int_0^t f_{01}(s)S_{12}(t-s)S_{02}(s)\,ds.\end{aligned} \tag{4}$$

In our package `smms`, we have implemented a general function giving the overall survival curve for any graph and models. When the times of entrance into absorbing states are observed exactly, the fitted curve in (4) may be compared to the non-parametric Kaplan–Meier estimate. This may indicate whether the parametric modelling assumptions appear to be reasonable, at least for the entire process from initial to absorbing states, see for example Figure 5 for an illustration.

The state occupancy probabilities $p_i(t)$ give the probability that a patient is found in state i at time t. These functions are referred to as the *prevalence* in parts of the literature[2]. Again, the exact expressions of these functions will depend on the multi-state graph, but for the three-state illness-death model from Figure 1 we have for example

$$\begin{aligned}
p_0(t) &= \Pr(T_{01} > t, T_{02} > t) = S_{01}(t)S_{02}(t) \\
p_1(t) &= \Pr(T_{01} + T_{12} > t, T_{01} < t, T_{01} < T_{02}) = \int_0^t f_{01}(s)S_{12}(t-s)S_{02}(s)\,ds \\
p_2(t) &= \Pr(T_{01} + T_{12} < t, T_{01} < t, T_{01} < T_{02}) + \Pr(T_{02} < t, T_{01} > T_{02}) \\
&= \int_0^t f_{01}(s)[1 - S_{12}(t-s)]S_{02}(s)\,ds + \int_0^t f_{02}(s)S_{01}(s)\,ds.
\end{aligned}$$

We have implemented a general function for the state occupancy probabilities in `smms`. The occupancy probabilities can be estimated non-parametrically using the Aalen–Johansen estimator[21]. This estimator does not rely on the Markov assumption[22], but requires that all transition times have been observed exactly. With interval censoring, this non-parametric estimator will be biased, but may still be compared to the fitted state occupancy curves to give an indication of potential lack of fit, see for example Figure 4.

Further, one is often interested in looking at the transition probabilities, namely given that a patient is found in state j at time v, what is the probability of occupying state i at a later time-point t: $P_{ij}(t, v) = Pr(Z(t) = i|Z(v) = j)$, with $Z(t)$ indicating the state at time t. For the three-state illness-death model from Figure 1 we can for example express the probability of being in state 2 at time t after having been in state 1 at time v,

$$\begin{aligned}
\Pr(Z(t) = 2|Z(v) = 1) &= \frac{\Pr(Z(t) = 2, Z(v) = 1)}{\Pr(Z(v) = 1)} = \frac{\Pr(v < T_{01} + T_{12} < t, T_{01} < v, T_{01} < T_{02})}{\Pr(T_{01} < v, T_{01} < T_{02})} \\
&= \frac{\int_0^v f_{01}(s)[S_{12}(v-s) - S_{12}(t-s)]S_{02}(s)\,ds}{\int_0^v f_{01}(s)S_{02}(s)\,ds},
\end{aligned}$$

and similarly for other pairs of states and multi-state graphs (we have a general function for this in `smms`).

After having fitted a multi-state model, our `smms` package provides pointwise approximate confidence bands for each of the quantities presented here using the delta method, see Section 2.4.

3. Computational considerations

Even though we provide a clear recipe for constructing the likelihood, as presented in Section 2.3, the steps may be somewhat involved and time-consuming to follow when faced with a new multi-state graph. The number of likelihood contribution types increases rapidly with the number of states in the graph, and the number of potential paths too. The efforts involved are too complex for many users of statistical methods, and we have therefore aimed at making our framework more widely available by implementing the recipe, and the maximisation of the likelihood, in an R package. The package consists of two main parts: (i) automatically writing out all the different likelihood contributions when faced with a given multi-state graph, (ii) efficiently evaluating and maximising the log-likelihood when faced with a graph and a dataset with states, time-points and covariates. The package is called `smms`, and the code and documentation is available on Github[23].

For part (i), we have implemented an efficient 'function factory' which provides the integrands, as R functions, for all possible paths in a given graph. We use the recipe from Section 2.3. The user is required to input the multi-state structure in the `igraph` format [24]; see vignettes on Github [23] for detailed explanations on the use of the `igraph` package, and other details on the use of `smms`.

Our package can currently handle any acyclic graph with up to six edges between the initial and absorbing state, but can easily be extended to larger graphs. Further, to make the code ready for part (ii), we have implemented a function which identifies the correct integral limits and other time-points necessary for the evaluation of the likelihood. The relevant time-points are the first and last time-points where each person is observed in a given transient state. For initial state(s) only the last time-point matters, for absorbing state(s) only the first. The main use of the integrands and time-points is the evaluation and maximisation of the likelihood in part (ii), but we additionally make use of the same functions for automatically writing out the full log-likelihood in latex format, as demonstrated for a variety of different models in the appendix.

In part (ii), we are optimising the log-likelihood to find the maximum likelihood estimates. The main challenge in this part is the computation of potentially high dimensional integrals. Currently, the package supports integrals up to dimension five, but may easily be extended. For the computations of integrals, we have chosen to use the `cubature` package in R [25], to benefit from their efficient C implementation of several numerical integration methods. In order to use this package, we have had to transform our formulas in order to obtain integrals with only numerical limits (as opposed to limits where variables are present in the inner integrals). This is achieved with a change-of-variable from the transition times to the cumulative transition times (for example $R = T_{01}$, $V = T_{01} + T_{12}$ and so on). This leads to transformations of the following type, for the likelihood contribution of a patient observed in state 0, 1 and 2 (type '012') in a four-state illness-death model (see Appendix A for the full likelihood)

$$\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} f_{01}(s) f_{12}(u) S_{23}(t_{2M} - u - s) S_{03}(s) S_{13}(u)\, du\, ds$$
$$= \int_{t_{0M}}^{t_{1m}} \int_{t_{1M}}^{t_{1m}} f_{01}(r) f_{12}(v-r) S_{23}(t_{2M} - v) S_{03}(r) S_{13}(v-r)\, dv\, dr.$$

In the `cubature` package the user can choose which numerical method to use and we mostly keep this option in our package. The exception is for integrals where the patient has travelled through transient states without being observed in them. For example, a patient may have travelled through states 0, 1, 2 in the three-state illness-death model, even though the patient is only observed in 0 and 2. Cases like these lead to two consecutive integrals over the same limit, but with constraints that one variable has to be larger than the other, and this is problematic for some of the numerical methods in the `cubature` package. For these integrals we have to use Monte Carlo methods, specifically the *vegas* option in the `cubature` package. An additional complication is that for one- and two dimensional integrals, the built-in R function `integrate` is actually faster than the numerical methods in the `cubature` package, and we therefore use that function for integrals of low dimension.

To minimise the computational time, we have included the possibility of parallel computations when evaluating the likelihood. The general idea is to use more than one processor to execute the computations [26]. We do this using the `mclapply` function in the R package `parallel`. Paralellisation is straightforward since the likelihood contribution of each patient in a dataset may be computed

independently of the rest of the patients. The `mclapply` function works on Unix-style operating systems via a fork mechanism. Therefore, parallelisation will not be available on Windows machines, and there the number of cores must be set to one. For more information about parallelisation in R and `mclapply`, see for example [26] and [27].

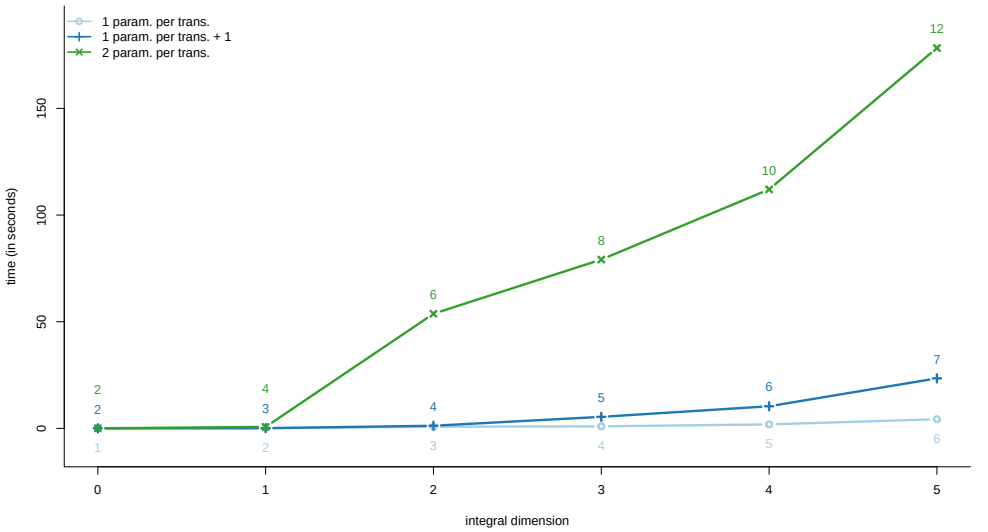


Figure 2. Time (in seconds) required to fit multi-state models for datasets of $n = 10$ patients as a function of integral dimension and number of parameters. Here we have not used the parallel computing option (so only a single core). We present three settings: one parameter per transition (using the exponential model for all transitions, light blue line), one parameter per transition plus one additional parameter (i.e. letting the first transition be Weibull and the rest exponentials, blue line), two parameters per transition (using Weibull for all transitions, green line). Coloured numbers indicate the total number of parameters to estimate in each setting.

The time and computational resources required to fit a model using the `smms` package will depend on many aspects of the multi-state graph and model. In general, the number of parameters to estimate and the maximal dimension of the integrals involved will be of prime importance, and we have striven to illustrate their effect in Figure 2. There, we record the time required to maximise the log-likelihood belonging to simulated datasets of $n = 10$ patients for progressive multi-state models with two to seven states, leading to one to six consecutive edges, meaning that the integral dimension ranges from zero to five. Progressive models are simple multi-state graphs where each state, except the initial state, only has a single possible transition into it (see for example Appendix C). For this example we do not use the parallel computing option. The light blue line gives the computational times for the case where we have one parameter per transition; here we use the exponential model for all transitions. The dark blue line gives the computational times for the case where we have one parameter per transition plus one additional parameter in total (giving two to seven parameters), meaning that we have a Weibull model for the first transition and exponential models for the remaining transitions. Finally, we also consider the case where we have two parameters per transition, and here we use the Weibull model for all transitions (green

line). The computational time increases slowly with integral dimension when the number of parameters increases by one for each additional edge in the graph (blue lines), but rather quickly when the number of parameters increases by two for each additional edge.

If one assumes that all transition times are exponentially distributed, one can find explicit expressions for all the integrals in the likelihood, see Section 4. We have currently not implemented these explicit solutions in our `smms` package, but as one can see from Figure 2 the optimisation with exponential distributions is reasonably fast anyway.

4. Connections to other multi-state frameworks

4.1. The time-homogeneous Markov model

The simplest parametric multi-state model is the homogeneous Markov model. That model relies on the Markov assumption, i.e. that the future evolution depends on the history only through the present state[28], and assumes that the transition intensity functions are time-independent, i.e. time-homogeneous. The homogeneous Markov model may be fit using the `msm` package[2]; the package also allows the user to relax the assumption of time-homogeneousness by letting the transition intensity functions be piecewise constant.

The framework presented in Section 2 constitutes a natural generalisation of the simple homogeneous Markov model: if one lets all T_{ij} be exponentially distributed, all the integrals in our likelihood construction simplify and one obtains the exact same likelihood as for the homogeneous Markov model (the one that is used in the `msm` package for example). This is expected from theory, see for instance Asanjarani et al.[15], and in the following we demonstrate the result in the context of the simple three-state illness-death model from Figure 1. With

$$T_{01} \sim \text{Expo}(\theta_0), \quad T_{12} \sim \text{Expo}(\theta_1), \quad \text{and} \quad T_{02} \sim \text{Expo}(\theta_2),$$

we obtain the following likelihood contributions using the formulas and notation from Section 2.2,

$$\begin{aligned}
L_0(\theta) &= e^{-(\theta_0+\theta_2)t_{0M}}, \\
L_{01}(\theta) &= \frac{\theta_0 e^{-\theta_1 t_{1M}}}{\theta_0+\theta_2-\theta_1}[e^{-(\theta_0+\theta_2-\theta_1)t_{0M}} - e^{-(\theta_0+\theta_2-\theta_1)t_{1m}}], \\
L_{02}(\theta) &= e^{-(\theta_0+\theta_2)t_{0M}} - e^{-(\theta_0+\theta_2)t_{2m}} - \frac{\theta_0 e^{-\theta_1 t_{2m}}}{\theta_0+\theta_2-\theta_1}[e^{-(\theta_0+\theta_2-\theta_1)t_{0M}} - e^{-(\theta_0+\theta_2-\theta_1)t_{2m}}], \\
L_{012}(\theta) &= \frac{\theta_0}{\theta_0+\theta_2-\theta_1}[e^{-\theta_1 t_{1M}} - e^{-\theta_1 t_{2m}}][e^{-(\theta_0+\theta_2-\theta_1)t_{0M}} - e^{-(\theta_0+\theta_2-\theta_1)t_{1m}}].
\end{aligned}$$

These are the same contributions one obtains using the method described by Jackson[29].

4.2. Two types of semi-Markov models

Semi-Markov models relax the Markov assumption by allowing the future evolution to depend on the time spent in the current state; see Asanjarani et al.[15] and Putter[28] for general reviews covering semi-Markov multi-state models. The most immediate specification of a semi-Markov model relies on defining an embedded Markov chain with jump probabilities π_{ij} (from state i to j), as well as probability distributions

for the *sojourn times* T^*_{ij}. We denote these variables with an asterisk to emphasise that they do not have the same interpretation, or role, as our T_{ij}. When a patient enters state i, its next state will be j with probability π_{ij}, and given that the next state is j, the time spent in state i is given by T^*_{ij} [15]. We will refer to this approach as the *embedded Markov chain* approach. Several recent contributions dealing with semi-Markov models and interval censoring make use of the embedded Markov chain approach [7,9,10]. If one assumes parametric models for the T^*_{ij}, the total number of parameters required with this approach is equal to the number of parameters governing the T^*_{ij} plus a number of π_{ij}. The number of free π_{ij} depends on the graph; naturally, the sum of all π_{ij} originating from the same state i must be one. In the example represented by Figure 1, there is only a single free π_{ij} since π_{12} must be equal to 1, and $\pi_{01} = 1 - \pi_{02}$.

Alternatively, one may specify a semi-Markov model as we have done in Section 2, by assuming latent T_{ij} for each transition and letting these T_{ij} compete with each other to determine the path of each patient. Most of the literature specify the distributions of the T_{ij} in terms of the intensity transition functions $h_{ij}(t)$, but as stated above we prefer to use the densities $f_{ij}(t)$ instead. We will refer to this approach as the *competing risk* approach, since this set-up actually amounts to a collection of competing risk models in each state [15]. Our impression is that, in the context of interval censoring, the competing risk approach has been less studied than the embedded Markov chain approach, despite its sparser parameterisation, which is particularly beneficial when dealing with the heavy computations required in the interval censoring case. The competing risk appoach is used by Chen et al. [30], Kang and Lagakos [6] and Kapetanakis et al. [8], which we return to in Section 7.

The two types of semi-Markov models are equivalent in the sense that one can always translate from one type to the other, even with covariates present. The formulas for the translation are provided by Asanjarani et al. [15]. Let us denote the density functions for the sojourn times T^*_{ij} in the embedded Markov chain approach by $g_{ij}(t;\gamma_{ij})$, with some transition specific parameters γ_{ij} (which we will omit in the following for ease of notation). We denote their cumulative distribution functions by $G_{ij}(t)$, and their associated hazard functions by $h^*_{ij}(t)$. We can *translate from the embedded Markov chain approach to the competing risk approach* by

$$S_{ij}(t) = \exp\Big\{ -\int_0^t h_{ij}(s)\,ds \Big\} = \exp\Big\{ -\int_0^t \frac{\pi_{ij} g_{ij}(s)}{\sum_{\ell\neq i}\pi_{i\ell}(1-G_{i\ell}(s))}\,ds \Big\}, \tag{5}$$

where the sum in the denominator is taken over all possible transitions going out of state i. If the only possible transition from state i is state j, the formula above simply implies that T_{ij} and T^*_{ij} will have the same distribution, but in general that will not be the case. We can *translate from the competing risk approach to the embedded Markov chain approach* by

$$\begin{aligned} \pi_{ij} &= \int_0^\infty f_{ij}(t) \prod_{\ell\neq(i,j)} S_{i\ell}(t)\,dt, \quad \text{and} \\ g_{ij}(t) &= \frac{f_{ij}(t)\prod_{\ell\neq(i,j)} S_{i\ell}(t)}{\pi_{ij}}, \end{aligned} \tag{6}$$

where the product of survival function is taken over all possible transitions going out of state i, other than the one to state j. As we clearly see here, in the competing risk approach the π_{ij} are not free, but functions

of the models we assume for the transition times. This approach will therefore have less parameters to estimate than the embedded Markov chain approach.

The formulas in Equations (5) and (6) clearly imply that if we have assumed named probability distributions in one approach, we will not necessarily get the same distributions in the other approach, or even some other named distributions. We will illustrate this (simple) fact, and the use of the formulas above, by a brief example making use of the three-state illness-death model from Figure 1. Let us first begin by specifying our model using the competing risk approach (as we do in the rest of this paper),

$$T_{01} \sim \text{Expo}(\theta_0), \quad T_{12} \sim \text{Expo}(\theta_1), \quad \text{and} \quad T_{02} \sim \text{Expo}(\theta_2).$$

We can use Equations (5) and (6) to find the equivalent embedded Markov chain model. For the transition between state 0 and 1 we get

$$\pi_{01} = \int_0^\infty f_{01}(t) S_{02}(t)\, dt = \theta_0/(\theta_0 + \theta_2),$$

$$g_{01}(t) = \frac{f_{01}(t) S_{02}(t)}{\pi_{01}} = (\theta_0 + \theta_2) \exp\{-(\theta_0 + \theta_2)t\}.$$

For the other transitions, we get $\pi_{12} = 1$ and $T^*_{12} \sim \text{Expo}(\theta_1)$ and $\pi_{22} = \theta_2/(\theta_0 + \theta_2)$ and $T^*_{02} \sim \text{Expo}(\theta_0 + \theta_2)$. So in this case, both approaches have exponential distributions for the transition/sojourn times, but with somewhat different parameters. Let us now specify the model using the embedded Markov chain approach,

$$T^*_{01} \sim \text{Expo}(\theta_0), \quad T^*_{12} \sim \text{Expo}(\theta_1), \quad \text{and} \quad T^*_{02} \sim \text{Expo}(\theta_2),$$

and with a single free π_{01} parameter (since $\pi_{12} = 1$, $\pi_{02} = 1 - \pi_{01}$). The implied model for the 0-1 transition from the competing risk approach is then

$$\begin{aligned} S_{01}(t) &= \exp\Big(- \int_0^t \frac{\pi_{01}\theta_0 \exp\{-\theta_0 s\}}{\pi_{01} \exp\{-\theta_0 s\} + (1 - \pi_{01}) \exp\{-\theta_2 s\}}\, ds \Big) \\ &= \exp\Big(- \frac{\theta_0}{\theta_2 - \theta_0} \log[\pi_{01} \exp\{(\theta_2 - \theta_1)t\} + 1 - \pi_{01}] \Big), \quad \text{for } \theta_0 \neq \theta_2. \end{aligned}$$

This survival function does not belong to any well-known named distribution. It is straightforward to see that in this case we only obtain exponential distributions for all transitions if we assume $\theta_0 = \theta_2$, in which case we have $T_{01} \sim \text{Expo}(\pi_{01}\theta_0)$ for example. Thus, it is only the competing risk approach that has the general property we describe in Section 4.1, that assuming exponential distributions for all transition times leads to the homogeneous Markov model. For the embedded Markov chain approach, this property requires the additional assumption that all sojourn times originating from the same state i have the same exponential parameter[15].

5. Illustration: monitoring heart transplant patients

We illustrate the use of our framework on the coronary allograft vasculopathy (CAV) dataset, one of the few openly available medical, panel datasets. The dataset was first used by Sharples et al.[31], and later as an

illustration by Jackson[2]. The patients in the dataset are heart transplant recipients and they are monitored for the post-transplant condition CAV. The time of heart transplantation constitutes a natural, common time origin for all patients; meaning that the time variable in the dataset indicates the time (in years) since heart transplantation. At each observation time-point the patients are assigned to one of four states: 'well' if they do not have CAV, 'mild CAV', 'severe CAV', and 'dead'. The time of death of the patients is recorded exactly, while the times of entrance into the CAV states are interval censored. Since CAV is generally considered an irreversible disease[2], we assume the four-state illness-death model in Figure 3 and discard the few observations belonging to patients which appear to move from more severe to less severe states. We end up with 2398 observations of 556 patients. Alternatively the discarded observations could have been treated as misclassifications, which would require an extension of our framework, see comments by Jackson[2].

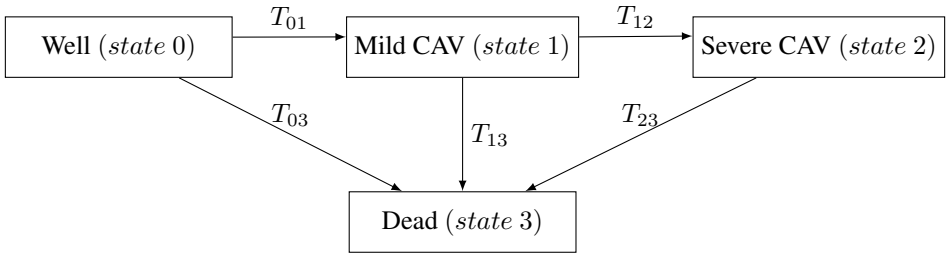


Figure 3. Multi-state graph for the CAV dataset - a four-state illness-death model

There are several potential covariates in the dataset, but in the following we have only made use of two, `dage` the age of the heart transplant donor (standardized), and `ihd` a binary covariate related to the reason for transplantation (whether the primary diagnosis was ischaemic heart disease or not). These choices are similar to the ones made by Jackson[2]; our results can therefore easily be compared. We include a detailed description of the use of our package for this dataset in a vignette[23] - for the model presented here and for a less complex model too.

One of the great advantages of our framework is the large flexibility available to the user in terms of specifying models for the transition times. Here we have chosen the following parametric models,

$$T_{01,k} \sim \text{Weibull}(\kappa_{01}, \lambda_{01,k}), \quad T_{12,k} \sim \text{Weibull}(\kappa_{12}, \lambda_{12,k}), \quad T_{23,k} \sim \text{Weibull}(\kappa_{23}, \lambda_{23,k}),$$
$$T_{03,k} \sim \text{GenWeibull}(\kappa_{03}, \lambda_{03,k}, \theta_{03}), \quad T_{13,k} \sim \text{Weibull}(\kappa_{13}, \lambda_{13}).$$

with $k = 1, \ldots, 556$ patients and the covariates influencing the scale parameter through $\lambda_{01,k} = \exp(\alpha_{01} + \beta_{1,01}\text{dage}_k + \beta_{2,01}\text{ihd}_k)$, and similarly for $\lambda_{12,k}$, $\lambda_{23,k}$ and $\lambda_{03,k}$. The generalised Weibull distribution is a one parameter extension of the ordinary Weibull distribution, see e.g.[7], with survival function

$$S_{ij}(t) = \exp\left[1 - \{1 + (t/\lambda_{ij})^{\kappa_{ij}}\}^{1/\theta_{ij}}\right].$$

When $\theta_{ij} = 1$ we are back to the ordinary Weibull (and when $\kappa_{ij} = 1$ too we have the exponential distribution). In total this gives us 19 parameters to estimate. The full log-likelihood for this case is given in Appendix A.

	Estimate	95% confidence interval
κ_{01}	1.52	$[1.35, 1.71]$
α_{01}	2.33	$[2.18, 2.48]$
$\beta_{1,01}$	-0.15	$[-0.24, -0.06]$
$\beta_{2,01}$	-0.32	$[-0.51, -0.13]$
κ_{12}	0.88	$[0.71, 1.10]$
α_{12}	0.98	$[0.61, 1.35]$
$\beta_{1,12}$	0.37	$[0.14, 0.60]$
$\beta_{2,12}$	-0.37	$[-0.83, 0.10]$
κ_{23}	0.65	$[0.50, 0.85]$
α_{23}	0.53	$[0.04, 1.02]$
$\beta_{1,23}$	0.38	$[0.03, 0.73]$
$\beta_{2,23}$	0.24	$[-0.30, 0.78]$

	Estimate	95% confidence interval
κ_{03}	0.63	$[0.41, 0.95]$
α_{03}	1.07	$[-2.11, 4.24]$
$\beta_{1,03}$	-1.26	$[-1.91, -0.61]$
$\beta_{2,03}$	-0.75	$[-2.04, 0.53]$
θ_{03}	8.03	$[2.10, 30.75]$
κ_{13}	8.00	$[3.31, 19.37]$
λ_{13}	6.35	$[5.45, 7.38]$

Table 1. Estimates and 95% approximate confidence intervals for the 19 parameters in the Weibull-GenWeibull model applied to the CAV dataset.

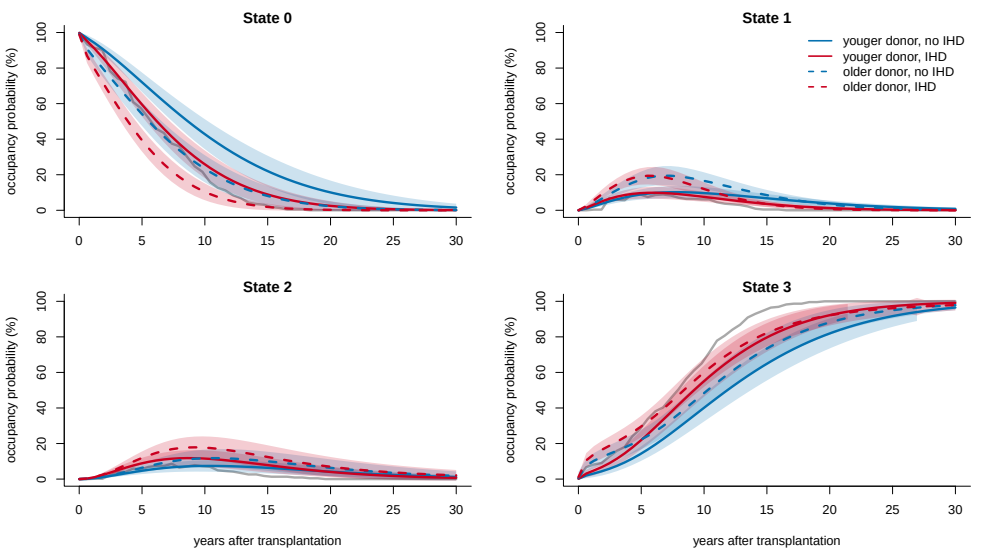


Figure 4. Coloured lines give the fitted occupancy probabilities for the four states (well, mild CAV, severe CAV, death), for patients without (blue) and with (red) IHD and for younger (17 years old, fully drawn) and older (around 40 years old, dashed) patients. The grey line are the non-parameteric estimates of the overall state occupancy probabilities. The shaded areas represent the pointwise (approximate) 95% confidence bands (obtained by the delta method).

We use our `smms` package to construct and optimise the log-likelihood. For this dataset, graph and model this takes around 50 minutes using 5 cores. To obtain approximate variances for all estimates

we need the observed Fisher information matrix (the hessian matrix from the optimisation) which takes another 20 minutes to compute. Estimates and approximate 95% confidence intervals are given in Table 1.

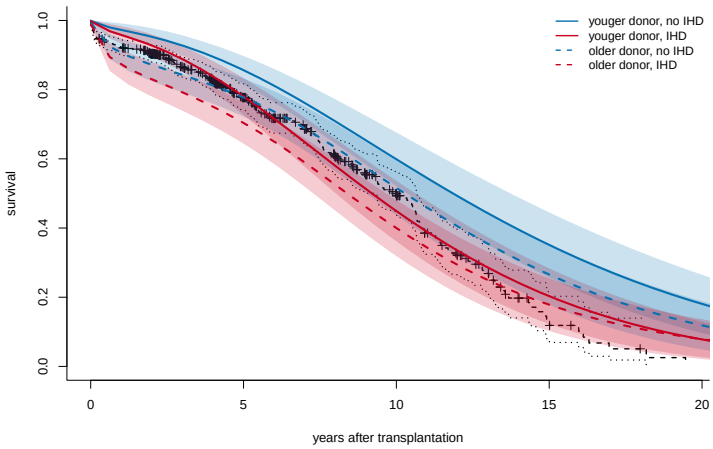


Figure 5. Coloured lines give the fitted overall survival curve for patients without (blue) and with (red) IHD and for younger (17 years old, fully drawn) and older (around 40 years old, dashed) patients. The black line is the non-parameteric estimates of the overall survival curve. The shaded areas represent the pointwise (approximate) 95% confidence bands (obtained by the delta method).

Older donor age appears to make the transition from well to mild CAV more likely, but appears to be protective for the further CAV progression (making the transition from mild to severe CAV, and from severe to death, less likely). A primary diagnosis of IHD is associated with a higher risk of developing CAV. High donor age and IHD are both associated with a higher risk of death without developing CAV. These effects are also apparent in the figure with state occupancy probabilities (Figure 4). Table 1 also indicates that most transition times have intensity transition functions that are far from constant, except for the transition between mild and severe CAV which could have been modelled by a simple exponential distribution instead. Further, this model has a far better fit to the data, as indicated from AIC, compared with the simple exponential model using the same covariates ($\mathrm{AIC} = -2786.5$ for the Weibull-GenWeibull model versus $\mathrm{AIC} = -2851.2$ for the exponential model).

Figures 4 and 5 may also be used as diagnostic plots as explained in Section 2.5. We see that the model fits reasonably well for all transitions, see vignette on GitHub[23] for similar figures for the exponential model.

6. Simulations

6.1. The degree of non-Markovness

Here we briefly investigate to what degree wrongly assuming a simple, homogeneous Markov model leads to biases in the estimated state occupancy probabilities, when the data are generated from a

semi-Markov model. Naturally, this will depend on the distance between the true model and the homogeneous Markov model, as we will see here. We consider a three-state illness-death model where the transition to the absorbing state is observed exactly, and in the first scenario we let the true model be

$$T_{01} \sim \text{Weibull}(5,3), \quad T_{12} \sim \text{Weibull}(3,4) \quad T_{02} \sim \text{Weibull}(2,4).$$

Then we fit a homogeneous Markov model as well as the true Weibull semi-Markov model (using the framework of this paper). The simulations are performed 100 times, in each simulation run we draw a dataset of 1000 patients. We estimate the state occupancy probabilities for each simulated dataset. In Figure 6, we present these state occupancy probabilities. We also include a thicker line indicating the state occupancy probabilities based on the average estimates from each model. As we would expect, the estimated state occupancies from the semi-Markov model are very close to the true state occupancy (grey line), while the Markov model produces state occupancies of widely different shapes. For example if we consider the plot for state 1, the true state occupancy probability is around 0.6 at time 4, while it is around 0.3 for the Markov model. In this scenario, the state occupancies are biased when using the Markov model instead of the semi-Markov model.

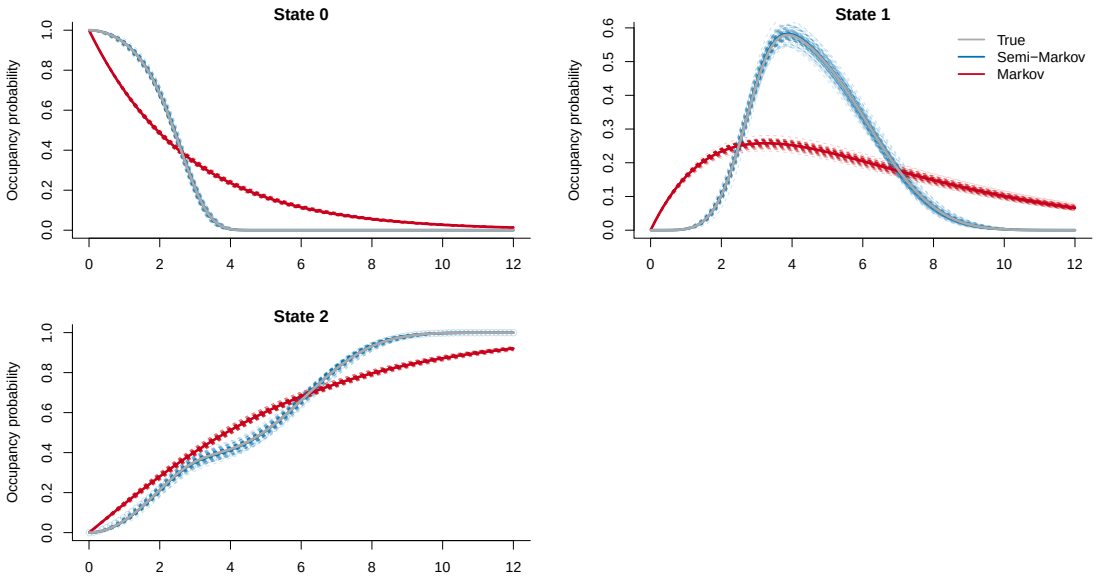


Figure 6. The dashed, coloured lines give the estimated state occupancy probabilities for the semi-Markov (blue) and Markov model (red) for a 100 simulated datasets, for the scenario where the true model is far away from the Markov model. The true state occupancy probabilities are given by the grey lines. The solid red and blue lines are the state occupancy probabilities using the mean of the estimated parameters.

In the next scenario we let the true model be close to a Markov model,

$$T_{01} \sim \text{Weibull}(1.1, 0.9), \quad T_{12} \sim \text{Weibull}(1, 1.1), \quad T_{02} \sim \text{Weibull}(0.95, 1.1).$$

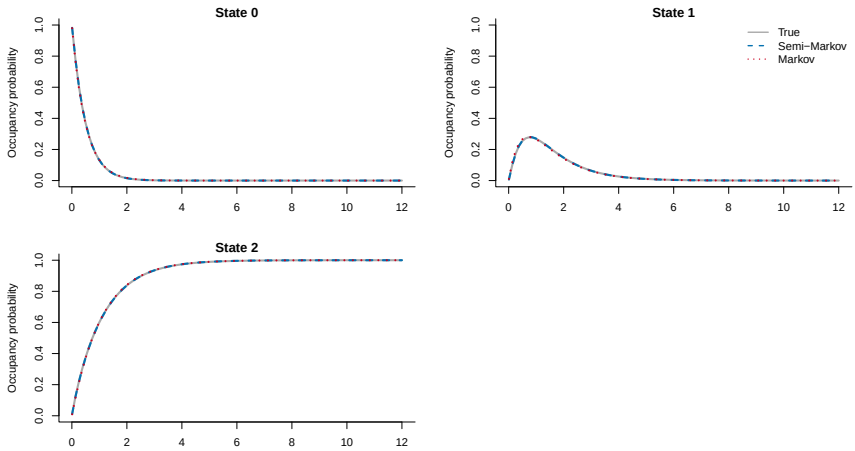


Figure 7. The dashed, coloured line gives the state occupancy probabilities using the mean of the estimated parameters, over a 100 simulated datasets, for the semi-Markov (blue) and Markov model (red), for the scenario where the true model is close to the Markov model. The true state occupancy probabilities are given by the grey lines.

These results are presented in Figure 7. In this scenario, the true model, semi-Markov model and Markov model are almost equal, and the state occupancies will have little bias when the Markov model is used instead of the semi-Markov model.

6.2. The relationship between the two semi-Markov approaches

Even though we can translate between the two semi-Markov approaches as we saw in Section 4, a choice of a named distributions in one approach will often lead to a non-standard distribution in the other. In practice, this means that at least one of the approaches will always be misspecified if we restrict ourselves to using named distributions. In this simulation study, we investigate to what degree using one or the other approach leads to misguided inference. Our main focus in this section will be the effect of covariates, and whether both approaches give similar inference for these effects.

We consider a three-state illness-death model with a single binary covariate which influences the transition times and the transition to the absorbing state is observed exactly. In the first scenario, we draw the data using the framework of this paper (i.e. the competing risk approach) with

$$T_{01,k} \sim \text{Weibull}(\kappa_{01,k}, \lambda_{01}), \quad T_{12,k} \sim \text{Weibull}(\kappa_{12,k}, \lambda_{12}), \quad T_{02,k} \sim \text{Weibull}(\kappa_{02,k}, \lambda_{02,k}), \quad (7)$$

where $k = 1, \ldots, 1000$ patients and the covariates influencing the scale parameter through $\kappa_{01,k} = \exp(\alpha_{01} + \beta_{01} x_k)$, and similarly for $\kappa_{12,k}$ and $\kappa_{02,k}$ (but for the last one $\beta_{02} = 0$). We run the simulations 100 times and fit the Weibull models using both the competing risks approach and the embedded Markov chain approach. We then know that the embedded Markov chain approach will be misspecified for all transitions, except for T_{12}, as we saw in Section 4. From Section 4 we also know

that some of the effect of the covariate will spill into π_{01}, and therefore we let $\pi_{01} = \exp(\alpha_{\pi,01} + \beta_{\pi,01}x_k)/\exp(\alpha_{\pi,01} + \beta_{\pi,01}x_k)$. The results are presented in Table 2. For many of the parameters, the average estimate in the competing risks approach and in the embedded Markov chain approach are relatively close to each other. The exceptions are for λ_{01} and λ_{02}, where the embedded Markov chain approach gives clearly different results. The relative biases for these parameters in the embedded Markov chain approach are also much higher than in the competing risk approach. The competing risk approach has generally slightly lower variances compared to the embedded Markov chain approach, as we would expect since it has two less parameters to estimate.

		Competing risks approach				Embedded Markov chain approach			
Parameter	True	Mean	Var.	Rel. bias	RMSE	Mean	Var.	Rel. bias	RMSE
$\alpha_{\pi,01}$	-	-	-	-	-	0.48	0.034	-	-
$\beta_{\pi,01}$	-	-	-	-	-	0.50	0.040	-	-
α_{01}	0.40	0.40	0.0035	-0.96	0.059	0.41	0.0047	1.33	0.068
β_{01}	-0.90	-0.90	0.0087	-0.27	0.093	-0.93	0.012	3.47	0.11
λ_{01}	1.82	1.83	0.010	0.19	0.10	1.26	0.0095	-30.78	0.57
α_{12}	1.60	1.61	0.019	0.71	0.14	1.61	0.020	0.81	0.14
β_{12}	-1.20	-1.21	0.026	0.72	0.16	-1.21	0.026	0.65	0.16
λ_{12}	0.61	0.60	0.00024	-0.33	0.016	0.61	0.00024	-0.14	0.016
α_{02}	0.20	0.20	0.0052	-0.89	0.072	0.22	0.0074	10.21	0.088
β_{02}	0.00	0.0086	0.0069	-	0.083	-0.048	0.016	-	0.14
λ_{02}	3.00	3.02	0.069	0.69	0.26	1.30	0.018	-56.84	1.71

Table 2. The true value, mean, variance, relative bias and RMSE for the parameters both for the competing risks and embedded Markov chain approach when we draw the data using the competing risks approach. The simulations are run 100 times for 1000 patients.

On average, both approaches give similar impressions concerning the effect of the covariate; β_{01} in the embedded Markov chain approach is a bit more negative than in the competing risk approach. The relative bias for this parameter is also relatively high in the embedded Markov chain approach. Also, β_{02} has a somewhat more negative effect according to the embedded Markov chain approach compared to the competing risks approach.

We can also compare the two approaches by studying plots of their estimated state occupancy probabilities, see Figure 8. Again, we see that the two approaches are quite close in this scenario. The embedded Markov chain displays some degree of bias, particularly for $x = 1$, and the plots also give the impression of more spread among the curves from the embedded Markov chain approach (in red) than among the curves from the competing risk approach (blue), which is consistent with the variances in Table 2.

In the second scenario, we draw the data using the embedded Markov chain approach with

$$T^*_{01,k} \sim \text{Weibull}(\kappa_{01,k}, \lambda_{01}), \quad T^*_{12,k} \sim \text{Weibull}(\kappa_{12,k}, \lambda_{12}), \quad T^*_{02,k} \sim \text{Weibull}(\kappa_{02,k}, \lambda_{02,k}),$$

where $k = 1, \ldots, 1000$ patients and the covariates influencing the scale parameter through $\kappa_{01,k} = \exp(\alpha_{01} + \beta_{01}x_k)$, and similarly for $\kappa_{12,k}$ and $\kappa_{02,k}$ (but for the last one $\beta_{02} = 0$). We also set the jump

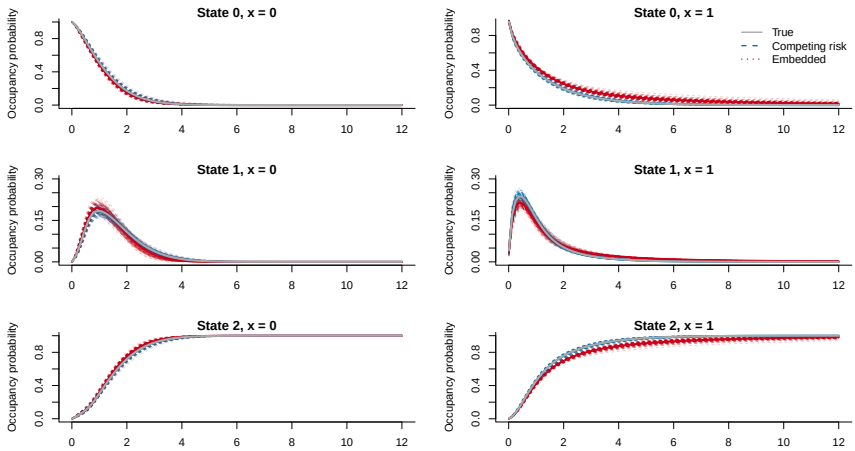


Figure 8. The dashed lines give the state occupancy probabilities using the estimated parameters for a 100 simulated datasets, for the competing risks approach (blue) and embedded Markov chain approach (red), for $x = 0$ and $x = 1$, when the competing risks approach is the true simulation model. The true state occupancy probabilities are given by the fully drawn, grey lines.

probability to $\pi_{01} \approx 0.5$. In this case, we know that the competing risk approach will be misspecified for all transitions, except for T_{12}, as we saw in Section 4. The results from this simulations study are presented in Table 3. In this scenario, there are clear differences between the estimates from the embedded Markov chain approach and the estimates from the competing risk approach for many of the parameters. The misspecified competing risk approach manages to find estimates of the covariate effect with the correct sign, but with a considerable bias especially for the 0-1 transition. We see the bias in the competing risk approach also in the state occupancy probability curves in Figure 9, particularly for State 1.

For these specific models and parameters, it appears that the embedded Markov chain approach was less biased, when misspecified (scenario 1), compared to the bias incurred by the competing risk approach when this was misspecified (scenario 2). In other words, it seems like the embedded Markov chain is more robust, when faced with the specific type of misspecification studied here, than the competing risk approach. This is likely related to the two additional parameters in the embedded Markov chain approach (scenario 1), which should allow for a more flexible fit to the data, even under misspecification. Whether this effect is true in general, also in the more realistic scenario where both approaches are misspecified, we leave as a topic for future research.

6.3. Coverage of approximate confidence intervals

As we mention in Section 2.4, we rely on general maximum likelihood theory for estimation and inference. Here we present a brief simulation study where we compute the realised coverage of the 95% approximate confidence intervals in a three-state illness-death model with Weibull transition times and

		Embedded Markov chain approach				Competing risks approach			
Parameter	True	Mean	Var.	Rel. bias	RMSE	Mean	Var.	Rel. bias	RMSE
π_{01}	0.50	0.50	0.0017	0.11	0.041	-	-	-	-
α_{01}	0.40	0.41	0.0074	1.62	0.086	0.16	0.0056	-60.44	0.25
β_{01}	-0.90	-0.90	0.011	0.24	0.11	-0.52	0.0058	-41.76	0.38
λ_{01}	1.82	1.85	0.047	1.37	0.22	4.56	0.35	150.37	2.80
α_{12}	1.60	1.63	0.027	1.91	0.17	1.64	0.026	2.67	0.17
β_{12}	-1.20	-1.22	0.038	1.54	0.20	-1.24	0.038	3.73	0.20
λ_{12}	0.61	0.60	0.00038	-0.35	0.019	0.61	0.00039	0.0079	0.020
α_{02}	0.20	0.21	0.0074	7.40	0.087	0.28	0.0056	38.30	0.11
β_{02}	0.00	0.0031	0.013	-	0.11	-0.044	0.0067	-	0.093
λ_{02}	3.00	2.98	0.13	-0.78	0.36	4.53	0.23	50.82	1.60

Table 3. The true value, mean, variance, relative bias and RMSE for the parameters both for the competing risks and embedded Markov chain approach when we draw the data using the embedded Markov chain approach. The simulations are run 100 times for 1000 patients.

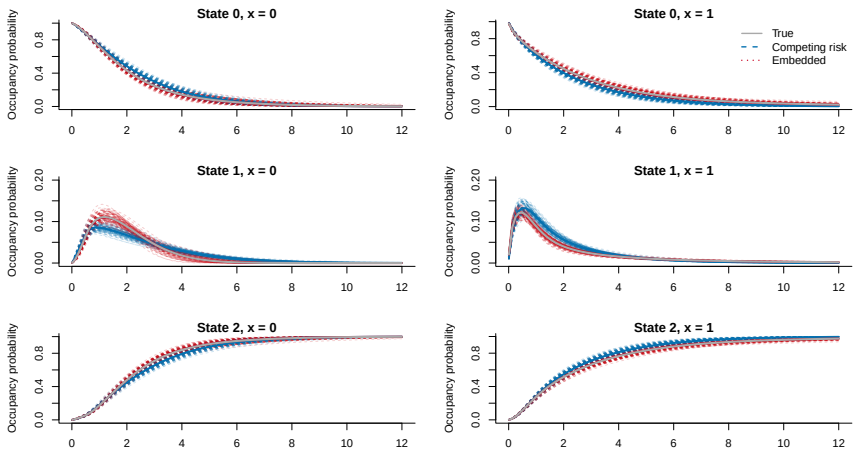


Figure 9. The dashed lines gives the state occupancy probabilities using the estimated parameters for a 100 simulated datasets, for the competing risks approach (blue) and embedded Markov chain approach (red), for $x = 0$ and $x = 1$, when the embedded Markov chain approach is the true simulation model. The true state occupancy probabilities are given by the fully drawn, grey lines.

a single binary covariate. We use the ordinary confidence interval construction assuming approximate normality, $\hat{\theta} \pm 1.96 \times \text{SE}$, where the standard error is obtained from the inverse of the observed Fisher information matrix (the hessian matrix from the optimisation). We simulate data from the model in (7) and investigate three scenarios with different numbers of patients $k = 200$, $k = 500$ and $k = 1000$. We simulate 500 datasets for each scenario. For each simulated dataset we compute 95% confidence intervals

for all nine parameters, and check whether these intervals cover the true values of the parameters. The realised coverage rates are given in Table 4 and we see that the coverage rates are close to their nominal value, even for a low sample size of 200.

Parameter	True	$k = 200$	$k = 500$	$k = 1000$
α_{01}	0.40	0.94	0.95	0.96
β_{01}	-0.90	0.95	0.95	0.95
λ_{01}	1.82	0.92	0.94	0.95
α_{12}	1.60	0.93	0.93	0.94
β_{12}	-1.20	0.92	0.93	0.95
λ_{12}	0.61	0.91	0.94	0.96
α_{02}	0.20	0.93	0.96	0.94
β_{02}	0.00	0.93	0.93	0.94
λ_{02}	3.00	0.92	0.94	0.95

Table 4. The true value and realised coverage rates for 95% confidence intervals for the parameters in a three-state illness-death model with Weibull transition times and one covariate, with three different sample sizes $k = 200$, $k = 500$ and $k = 1000$. For each scenario the simulation are run 500 times.

7. Discussion and concluding remarks

We have presented a general framework for parametric inference with interval censored multi-state data. Semi-Markov models allow for much more flexible, and realistic, multi-state modelling than the simple, time-homogeneous Markov model, for example allowing for intensity functions with a range of different shapes. There are few computational tools for fitting semi-Markov models to panel data, and we have therefore implemented our general framework in an R package, which can accommodate any parametric model for the transition times and include covariates in various ways.

We have spent some efforts in clarifying the relationship between our framework and existing methodologies, particularly existing semi-Markov frameworks. In that context we have distinguished between what we refer to as the embedded Markov chain approach, and the competing risk approach, within which our framework is found. Chen et al.[30], Kapetanakis et al.[8] and Kang and Lagakos[6] have also investigated models which belong to the competing risk approach, and are therefore closely related to our framework. Chen et al.[30] and Kapetanakis et al.[8] consider specific multi-state structures: Chen et al.[30] focus on simple progressive models, while Kapetanakis et al.[8] consider a three-state illness-death model. Kang and Lagakos[6] present a more general framework, but require that the transition intensities from at least one of the states are time-homogeneous. This allows them to handle multi-state graphs with cycles, which we come back to below.

An assumption in our model framework that may be criticised is the assumption of independent transition times. This is a strong assumption, which for some applications may not be realistic, but we can to some extent get around this assumption by including covariates. Our framework can be extended to handle dependent transition times, but the results will likely be highly dependent on the form of dependency structure chosen, and these assumptions would be largely untestable since the transition times are latent. Similarly to other multi-state frameworks, we also assume that the graph is a correct representation of the process. If the graph is wrong, the subsequent inference will be wrong too. This has

to some extent been discussed in Jepsen et al.[13], where they study a disease called cirrhosis and discuss how it is important for clinicians to construct these studies in a way that can be studied using multi-state models.

In this paper, we do not cover multi-state models with cyclic graphs; often referred to as models with back transitions in the literature. Cyclic graphs pose serious challenges for inference with panel data. The problem with interval censoring and cyclic graphs is that when you observe a path, for example '01', you may have infinitely many underlying sample paths; i.e. the patient may have travelled infinitely many times back and forth between states 0 and 1. In the literature, this problem has been addressed by simplifying assumptions or particular computational tools. The aforementioned paper by Kang and Lagakos[6] obtain a tractable likelihood by assuming that the transition intensities from at least one of the states are time-homogeneous. Alternatively, Wei and Kryscio[9] circumvented the problem by assuming that their dataset contains no unobserved states, i.e. that the sequence of states observed is equal to the sequence of visited states. Both of these approaches may be considered if one seeks to extend our framework to cover cyclic graphs, but none of these options is completely unproblematic: letting the transition intensities from one of the states be time-homogeneous may be unsatisfactory if this assumption does not fit well with the data, and assuming that the dataset contains no unobserved states may lead to misleading inference unless the patients are observed very frequently.

Some recent papers address the problem of cyclic graphs with panel data by applying missing data techniques. The idea is based on the fact, which we state in Section 2, that if all transition times had been observed exactly the likelihood would be a simple product of densities and survival functions, without any integrals. One can make use of this simple likelihood by sampling a number of likely realisations of paths and transition times given the observed data and the model, and then proceed with inference by stochastic expectation-maximization[10] or Bayesian methods[32,33]. Aralis and Brookmeyer[10] sample trajectories by rejection sampling, while Barone and Tancredi[32] use Markov Chain Monte Carlo. The computational tools used in those papers may naturally also be employed in the acyclic case, and it would be an interesting topic for future research to compare those tools with our framework, both with respect to estimation quality and computational time.

Above, and in our `smms` package, we have used the general log-likelihood development in order to carry out ML estimation, inference, and model selection. One may also use these log-likelihood formulae to carry out Bayesian estimation and inference, however. In situations with clear prior information, for some of the components of the model, coupled perhaps with noninformative priors for other components, posterior distributions may be found via MCMC algorithms. Suitable extensions of our methods can then be used in setups as for the dementia progression analyses of Williams et al.[11]. In Williams et al.[11], they study models with a hidden layer. This is also a possible extension for our current model framework. The start ideas for doing this are discussed by Aastveit[16]. Another potential extension is to include the possibility of misclassification of states[6,9].

When applying our multi-state setup there will typically be several candidate models, particularly when it comes to modelling the influence of covariates. In our applications we have been guided by the AIC model selection criterion, as the relevant theory applies to our classes of log-likelihood functions. Also versions of the focused information criterion FIC[20,34] may be developed here, suitable for occasions where there is a primary quantity of interest, for which the most accurate estimation is required. One may also develop extensions of goodness-of-fit monitoring processes, see[35], for multi-state models, but this is outside the scope of the present article.

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

Acknowledgements

This paper was revised as part of the 'Stability and Change' project at the Centre for Advanced Study at The Norwegian Academy of Science and Letters. We are grateful for the constructive comments from the anonymous reviewer.

Supplemental material

All R code used for the preparation of this article is available on GitHub[23]. The data used in the application section is available from the `msm` package[2].

References

1. Kalbfleisch J and Lawless JF. The analysis of panel data under a Markov assumption. *J Am Stat Assoc* 1985; **80**: 863–871.
2. Jackson CH. Multi-state models for panel data: the msm package for R. *J Stat Softw* 2011; **38**: 1–28.
3. de Wreede LC, Fiocco M and Putter H. mstate: an R package for the analysis of competing risks and multi-state models. *J Stat Softw* 2011; **38**: 1–30.
4. Król A and Saint-Pierre P. SemiMarkov: An R package for parametric estimation in multi-state semi-Markov models. *J Stat Softw* 2015; **66**: 1–16.
5. Jackson CH. flexsurv: a platform for parametric survival modeling in R. *J Stat Softw* 2016; **70**: 1–38.
6. Kang M and Lagakos SW. Statistical methods for panel data from a semi-Markov process, with application to HPV. *Biostatistics* 2006; **8**: 252–264.
7. Foucher Y, Giral M, Soulillou J et al. A flexible semi-Markov model for interval-censored data and goodness-of-fit testing. *Stat Methods Med Res* 2010; **19**: 127–145.
8. Kapetanakis V, Matthews FE and van den Hout A. A semi-Markov model for stroke with piecewise-constant hazards in the presence of left, right and interval censoring. *Stat Med* 2013; **32**: 697–713.
9. Wei S and Kryscio RJ. Semi-Markov models for interval censored transient cognitive states with back transitions and a competing risk. *Stat Methods Med Res* 2016; **25**: 2909–2924.
10. Aralis H and Brookmeyer R. A stochastic estimation procedure for intermittently-observed semi-Markov multistate models with back transitions. *Stat Methods Med Res* 2019; **28**: 770–787.
11. Williams JP, Storlie CB, Therneau TM et al. A Bayesian approach to multistate hidden Markov models: application to dementia progression. *J Am Stat Assoc* 2020; **115**: 16–31.
12. Gentleman R, Lawless J, Lindsey J et al. Multi-state Markov models for analysing incomplete disease history data with illustrations for HIV disease. *Stat Med* 1994; **13**: 805–821.
13. Jepsen P, Vilstrup H and Andersen PK. The clinical course of cirrhosis: the importance of multistate models and competing risks analysis. *Hepatology* 2015; **62**: 292–302.
14. Commenges D. Inference for multi-state models from interval-censored data. *Stat Methods Med Res* 2002; **11**: 167–182.
15. Asanjarani A, Liquet B and Nazarathy Y. Estimation of semi-Markov multi-state models: a comparison of the sojourn times and transition intensities approaches. *Int J Biostat* 2021; **18**: 1–20.

16. Aastveit ME. *Multi-State Models for Interval-Censored Data with Transition Times from Gamma Processes*. Master's Thesis, 2021.
17. Hougaard P. Multi-state models: a review. *Lifetime Data Anal* 1999; **5**: 239–264.
18. Huang J and Wellner JA. Interval censored survival data: a review of recent progress. In *Proceedings of the first Seattle symposium in Biostatistics*. Springer, 1997, pp. 123–169.
19. Kim JS. Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *J Royal Stat Soc: Ser B (Stat Methodol)* 2003; **65**: 489–502.
20. Claeskens G and Hjort NL. *Model Selection and Model Averaging*. Cambridge: Cambridge University Press, 2008.
21. Aalen O, Borgan Ø and Gjessing H. *Survival and event history analysis: a process point of view*. Springer Science & Business Media, 2008.
22. 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.
23. Aastveit M and Cunen C. smms: Semi-Markov Multi-State models for interval censored data. 2022. `https://github.com/NorskRegnesentral/smms/`.
24. Csardi G and Nepusz T. The igraph software package for complex network research. *Int J Complex Syst* 2006; **1695**: 1–9. URL `https://igraph.org`.
25. Narasimhan B, Johnson SG, Hahn T et al. *cubature: Adaptive Multivariate Integration over Hypercubes*, 2022. URL `https://bnaras.github.io/cubature/`. R package version 2.0.4.4.
26. Peng RD. *R Programming for Data Science*. Leanpub Victoria, BC, Canada, 2016.
27. R Core Team. *Package 'parallel'*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL `https://stat.ethz.ch/R-manual/R-devel/library/parallel/doc/parallel.pdf`.
28. Putter H, Fiocco M and Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* 2007; **26**: 2389–2430.
29. Jackson C. Multi-state modelling with R: the msm package. *Cambridge, UK* 2007; : 1–53.
30. Chen THH, Yen MF, Shiu MN et al. Stochastic model for non-standard case-cohort design. *Stat Med* 2004; **23**: 633–647.
31. Sharples LD, Jackson CH, Parameshwar J et al. Diagnostic accuracy of coronary angiography and risk factors for post–heart-transplant cardiac allograft vasculopathy. *Transplantation* 2003; **76**: 679–682.
32. Barone R. and Tancredi A. Bayesian inference for discretely observed continuous time multi-state models. *stat Med* 2022: 1–15.
33. Tancredi A. Approximate Bayesian inference for discretely observed continuous-time multi-state models. *Biometrics* 2019; **75**: 966–977.
34. Cunen C, Walløe L and Hjort NL. Focused model selection for linear mixed models with an application to whale ecology. *Ann Appl Stat* 2020; **14**: 872 – 904.
35. Hjort NL. Goodness of fit tests in models for life history data based on cumulative hazard rates. *Ann Stat* 1990; **18**: 1221–1258.

Appendix

Here we display the log-likelihoods belonging to a selection of multi-state graphs. These formulae have been automatically generated by our `smms` package. For most graphs we present two likelihoods: one for the case where all transition times are censored, and another for the common case where the transition time into the absorbing state(s) is observed exactly. For each multi-state graphs a number of likelihood contribution types may be observed (belonging to different combinations of states), n_0 denotes the number of patients observed only in state 0 for example, and n_{01} denotes the number of patients observed in state 0 and in state 1, and so on. θ indicates the full parameter vector. The observation times pertaining to state i are denoted by t_{im} or t_{iM}, with m indicating the *minimal* or first observation time in state i (for a given patient), and M indicating the *maximal* or last observation time in state i. Sometimes we will have $t_{im} = t_{iM}$, but this does not require any changes to the formulae below. When the time of entry into the absorbing state is observed exactly we denote that time-point by t_{im} (even though there are no more observations in state i in that case).

A. Four-state illness-death model

When the entry into the absorbing state is interval censored too:

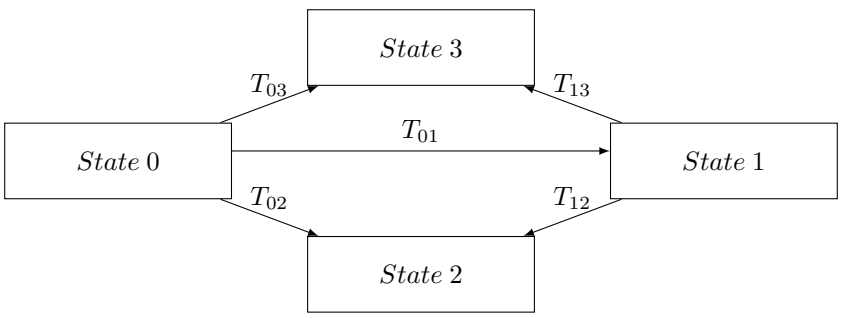
$$
\begin{aligned}
\ell_n(\theta) = &\sum_{k=0}^{n_0} \log\{S_{01}(t_{0M})S_{03}(t_{0M})\} + \sum_{k=0}^{n_{01}} \log\{\int_{t_{0M}}^{t_{1m}} f_{01}(s)S_{03}(s)S_{12}(t_{1M}-s)S_{13}(t_{1M}-s)\,ds\}+\\
&\sum_{k=0}^{n_{012}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} f_{01}(s)f_{12}(u)S_{03}(s)S_{13}(u)S_{23}(t_{2M}-s-u)\,du\,ds\}+\\
&\sum_{k=0}^{n_{02}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} f_{01}(s)f_{12}(u)S_{03}(s)S_{13}(u)S_{23}(t_{2M}-s-u)\,du\,ds\}+\\
&\sum_{k=0}^{n_{0123}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} f_{01}(s)f_{12}(u)(S_{23}(t_{2M}-s-u)-S_{23}(t_{3m}-s-u))S_{03}(s)S_{13}(u)\,du\,ds\}+\\
&\sum_{k=0}^{n_{03}} \log\{\int_{t_{0M}}^{t_{3m}} \int_{0}^{t_{3m}-s} f_{01}(s)f_{12}(u)(1-S_{23}(t_{3m}-s-u))S_{03}(s)S_{13}(u)\,du\,ds+\\
&\int_{t_{0M}}^{t_{3m}} \int_{0}^{t_{3m}-s} f_{01}(s)f_{13}(u)S_{03}(s)S_{12}(u)\,du\,ds + \int_{t_{0M}}^{t_{3m}} f_{03}(s)S_{01}(s)\,ds\}+\\
&\sum_{k=0}^{n_{013}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{3m}-s} f_{01}(s)f_{12}(u)(1-S_{23}(t_{3m}-s-u))S_{03}(s)S_{13}(u)\,du\,ds+\\
&\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{3m}-s} f_{01}(s)f_{13}(u)S_{03}(s)S_{12}(u)\,du\,ds\}+\\
&\sum_{k=0}^{n_{023}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} f_{01}(s)f_{12}(u)(S_{23}(t_{2M}-s-u)-S_{23}(t_{3m}-s-u))S_{03}(s)S_{13}(u)\,du\,ds\}.
\end{aligned}
$$

When the entry into the absorbing state is observed exactly:

$$\ell_n(\theta) = \sum_{k=0}^{n_0} \log\{S_{01}(t_{0M})S_{03}(t_{0M})\} + \sum_{k=0}^{n_{01}} \log\{\int_{t_{0M}}^{t_{1m}} f_{01}(s)S_{03}(s)S_{12}(t_{1M}-s)S_{13}(t_{1M}-s)\,ds\}+$$
$$\sum_{k=0}^{n_{012}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} f_{01}(s)f_{12}(u)S_{03}(s)S_{13}(u)S_{23}(t_{2M}-s-u)\,du\,ds\}+$$
$$\sum_{k=0}^{n_{02}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} f_{01}(s)f_{12}(u)S_{03}(s)S_{13}(u)S_{23}(t_{2M}-s-u)\,du\,ds\}+$$
$$\sum_{k=0}^{n_{0123}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} f_{01}(s)f_{12}(u)f_{23}(t_{3m}-s-u)S_{03}(s)S_{13}(u)\,du\,ds\}+$$
$$\sum_{k=0}^{n_{03}} \log\{\int_{t_{0M}}^{t_{3m}} \int_{0}^{t_{3m}-s} f_{01}(s)f_{12}(u)f_{23}(t_{3m}-s-u)S_{03}(s)S_{13}(u)\,du\,ds+$$
$$\int_{t_{0M}}^{t_{3m}} f_{01}(s)f_{13}(t_{3m}-s)S_{03}(s)S_{12}(t_{3m}-s)\,ds + f_{03}(t_{3m})S_{01}(t_{3m})\}+$$
$$\sum_{k=0}^{n_{013}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{3m}-s} f_{01}(s)f_{12}(u)f_{23}(t_{3m}-s-u)S_{03}(s)S_{13}(u)\,du\,ds+$$
$$\int_{t_{0M}}^{t_{1m}} f_{01}(s)f_{13}(t_{3m}-s)S_{03}(s)S_{12}(t_{3m}-s)\,ds\}+$$
$$\sum_{k=0}^{n_{023}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} f_{01}(s)f_{12}(u)f_{23}(t_{3m}-s-u)S_{03}(s)S_{13}(u)\,du\,ds\}.$$

This is the log-likelihood for the CAV application.

B. Multi-state model from Foucher et al.[7]



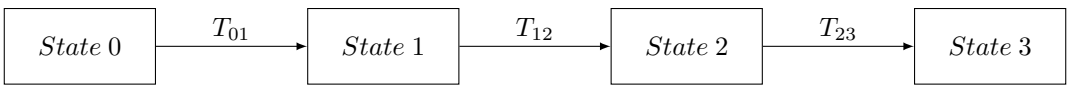
When the entry into the absorbing state is interval censored too:

$$\begin{aligned}\ell_n(\theta) = &\sum_{k=0}^{n_0} \log\{S_{01}(t_{0M})S_{02}(t_{0M})S_{03}(t_{0M})\} + \sum_{k=0}^{n_{01}} \log\{\int_{t_{0M}}^{t_{1m}} f_{01}(s)S_{02}(s)S_{03}(s)S_{12}(t_{1M}-s)S_{13}(t_{1M}-s)\,ds\}+\\ &\sum_{k=0}^{n_{012}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} f_{01}(s)f_{12}(u)S_{02}(s)S_{03}(s)S_{13}(u)\,du\,ds\}+\\ &\sum_{k=0}^{n_{02}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} f_{01}(s)f_{12}(u)S_{02}(s)S_{03}(s)S_{13}(u)\,du\,ds + \int_{t_{0M}}^{t_{2m}} f_{02}(s)S_{01}(s)S_{03}(s)\,ds\}+\\ &\sum_{k=0}^{n_{013}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{3m}-s} f_{01}(s)f_{13}(u)S_{02}(s)S_{03}(s)S_{12}(u)\,du\,ds\}+\\ &\sum_{k=0}^{n_{03}} \log\{\int_{t_{0M}}^{t_{3m}} \int_{0}^{t_{3m}-s} f_{01}(s)f_{13}(u)S_{02}(s)S_{03}(s)S_{12}(u)\,du\,ds + \int_{t_{0M}}^{t_{3m}} f_{03}(s)S_{01}(s)S_{02}(s)\,ds\}.\end{aligned}$$

When the entry into the absorbing state is observed exactly (this is the situation in Foucher et al.[7]):

$$\begin{aligned}\ell_n(\theta) = &\sum_{k=0}^{n_0} \log\{S_{01}(t_{0M})S_{02}(t_{0M})S_{03}(t_{0M})\} + \sum_{k=0}^{n_{01}} \log\{\int_{t_{0M}}^{t_{1m}} f_{01}(s)S_{02}(s)S_{03}(s)S_{12}(t_{1M}-s)S_{13}(t_{1M}-s)\,ds\}+\\ &\sum_{k=0}^{n_{012}} \log\{\int_{t_{0M}}^{t_{1m}} f_{01}(s)f_{12}(t_{2m}-s)S_{02}(s)S_{03}(s)S_{13}(t_{2m}-s)\,ds\}+\\ &\sum_{k=0}^{n_{02}} \log\{\int_{t_{0M}}^{t_{2m}} f_{01}(s)f_{12}(t_{2m}-s)S_{02}(s)S_{03}(s)S_{13}(t_{2m}-s)\,ds + f_{02}(t_{2m})S_{01}(t_{2m})S_{03}(t_{2m})\}+\\ &\sum_{k=0}^{n_{013}} \log\{\int_{t_{0M}}^{t_{1m}} f_{01}(s)f_{13}(t_{3m}-s)S_{02}(s)S_{03}(s)S_{12}(t_{3m}-s)\,ds\}+\\ &\sum_{k=0}^{n_{03}} \log\{\int_{t_{0M}}^{t_{3m}} f_{01}(s)f_{13}(t_{3m}-s)S_{02}(s)S_{03}(s)S_{12}(t_{3m}-s)\,ds + f_{03}(t_{3m})S_{01}(t_{3m})S_{02}(t_{3m})\}.\end{aligned}$$

C. Four-state progressive model



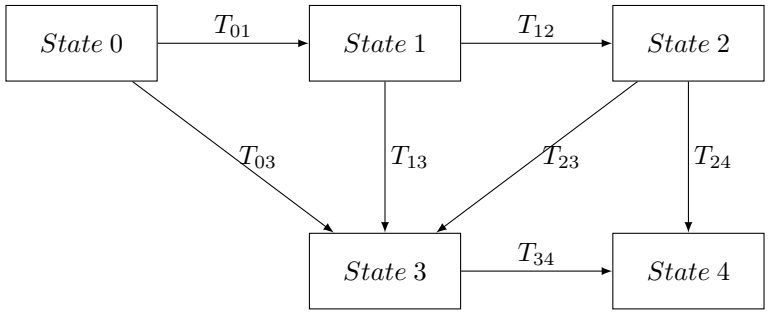
When the entry into the absorbing state is interval censored too:

$$\begin{aligned}\ell_n(\theta) = &\sum_{k=0}^{n_0} \log\{S_{01}(t_{0M})\} + \sum_{k=0}^{n_{01}} \log\{\int_{t_{0M}}^{t_{1m}} f_{01}(s) S_{12}(t_{1M}-s)\, ds\} + \\ &\sum_{k=0}^{n_{012}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} f_{01}(s) f_{12}(u) S_{23}(t_{2M}-s-u)\, du\, ds\} + \\ &\sum_{k=0}^{n_{02}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} f_{01}(s) f_{12}(u) S_{23}(t_{2M}-s-u)\, du\, ds\} + \\ &\sum_{k=0}^{n_{0123}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} f_{01}(s) f_{12}(u) (S_{23}(t_{2M}-s-u) - S_{23}(t_{3m}-s-u))\, du\, ds\} + \\ &\sum_{k=0}^{n_{03}} \log\{\int_{t_{0M}}^{t_{3m}} \int_{0}^{t_{3m}-s} f_{01}(s) f_{12}(u) (1 - S_{23}(t_{3m}-s-u))\, du\, ds\} + \\ &\sum_{k=0}^{n_{013}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{3m}-s} f_{01}(s) f_{12}(u) (1 - S_{23}(t_{3m}-s-u))\, du\, ds\} + \\ &\sum_{k=0}^{n_{023}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} f_{01}(s) f_{12}(u) (S_{23}(t_{2M}-s-u) - S_{23}(t_{3m}-s-u))\, du\, ds\}.\end{aligned}$$

When the entry into the absorbing state is observed exactly:

$$\begin{aligned}\ell_n(\theta) = &\sum_{k=0}^{n_0} \log\{S_{01}(t_{0M})\} + \sum_{k=0}^{n_{01}} \log\{\int_{t_{0M}}^{t_{1m}} f_{01}(s) S_{12}(t_{1M}-s)\, ds\} + \\ &\sum_{k=0}^{n_{012}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} f_{01}(s) f_{12}(u) S_{23}(t_{2M}-s-u)\, du\, ds\} + \\ &\sum_{k=0}^{n_{02}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} f_{01}(s) f_{12}(u) S_{23}(t_{2M}-s-u)\, du\, ds\} + \\ &\sum_{k=0}^{n_{0123}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} f_{01}(s) f_{12}(u) f_{23}(t_{3m}-s-u)\, du\, ds\} + \\ &\sum_{k=0}^{n_{03}} \log\{\int_{t_{0M}}^{t_{3m}} \int_{0}^{t_{3m}-s} f_{01}(s) f_{12}(u) f_{23}(t_{3m}-s-u)\, du\, ds\} + \\ &\sum_{k=0}^{n_{013}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{3m}-s} f_{01}(s) f_{12}(u) f_{23}(t_{3m}-s-u)\, du\, ds\} + \\ &\sum_{k=0}^{n_{023}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} f_{01}(s) f_{12}(u) f_{23}(t_{3m}-s-u)\, du\, ds\}.\end{aligned}$$

D. A five-state model



$$\ell_n(\theta) = \sum_{k=0}^{n_0} \log\{S_{01}(t_{0M})S_{03}(t_{0M})\} + \sum_{k=0}^{n_{01}} \log\{\int_{t_{0M}}^{t_{1m}} f_{01}(s)S_{03}(s)S_{12}(t_{1M}-s)S_{13}(t_{1M}-s)\,ds\}+$$

$$\sum_{k=0}^{n_{012}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} f_{01}(s)f_{12}(u)S_{03}(s)S_{13}(u)S_{23}(t_{2M}-s-u)S_{24}(t_{2M}-s-u)\,du\,ds\}+$$

$$\sum_{k=0}^{n_{02}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} f_{01}(s)f_{12}(u)S_{03}(s)S_{13}(u)S_{23}(t_{2M}-s-u)S_{24}(t_{2M}-s-u)\,du\,ds\}+$$

$$\sum_{k=0}^{n_{0123}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} \int_{t_{2M}-s-u}^{t_{3m}-s-u} f_{01}(s)f_{12}(u)f_{23}(r)S_{03}(s)S_{13}(u)S_{24}(r)S_{34}(t_{3M}-s-u-r)\,dr\,du\,ds\}+$$

$$\sum_{k=0}^{n_{03}} \log\{\int_{t_{0M}}^{t_{3m}} \int_{0}^{t_{3m}-s} \int_{0}^{t_{3m}-s-u} f_{01}(s)f_{12}(u)f_{23}(r)S_{03}(s)S_{13}(u)S_{24}(r)S_{34}(t_{3M}-s-u-r)\,dr\,du\,ds+$$

$$\int_{t_{0M}}^{t_{3m}} \int_{0}^{t_{3m}-s} f_{01}(s)f_{13}(u)S_{03}(s)S_{12}(u)S_{34}(t_{3M}-s-u)\,du\,ds + \int_{t_{0M}}^{t_{3m}} f_{03}(s)S_{01}(s)S_{34}(t_{3M}-s)\,ds\}+$$

$$\sum_{k=0}^{n_{013}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{3m}-s} \int_{0}^{t_{3m}-s-u} f_{01}(s)f_{12}(u)f_{23}(r)S_{03}(s)S_{13}(u)S_{24}(r)S_{34}(t_{3M}-s-u-r)\,dr\,du\,ds+$$

$$\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{3m}-s} f_{01}(s)f_{13}(u)S_{03}(s)S_{12}(u)S_{34}(t_{3M}-s-u)\,du\,ds\}+$$

$$\sum_{k=0}^{n_{023}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} \int_{t_{2M}-s-u}^{t_{3m}-s-u} f_{01}(s)f_{12}(u)f_{23}(r)S_{03}(s)S_{13}(u)S_{24}(r)S_{34}(t_{3M}-s-u-r)\,dr\,du\,ds\}+$$

$$\sum_{k=0}^{n_{01234}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} \int_{t_{2M}-s-u}^{t_{3m}-s-u} f_{01}(s)f_{12}(u)f_{23}(r)(S_{34}(t_{3M}-s-u-r)-S_{34}(t_{4m}-s-u-r))S_{03}(s)S_{13}(u)S_{24}(r)\,dr\,du\,ds\}+$$

$$\sum_{k=0}^{n_{04}} \log\{\int_{t_{0M}}^{t_{4m}} \int_{0}^{t_{4m}-s} \int_{0}^{t_{4m}-s-u} f_{01}(s)f_{12}(u)f_{23}(r)(1-S_{34}(t_{4m}-s-u-r))S_{03}(s)S_{13}(u)S_{24}(r)\,dr\,du\,ds+$$

$$\int_{t_{0M}}^{t_{4m}} \int_{0}^{t_{4m}-s} \int_{0}^{t_{4m}-s-u} f_{01}(s)f_{12}(u)f_{24}(r)S_{03}(s)S_{13}(u)S_{23}(r)\,dr\,du\,ds+$$

$$\int_{t_{0M}}^{t_{4m}} \int_{0}^{t_{4m}-s} f_{01}(s)f_{13}(u)(1-S_{34}(t_{4m}-s-u))S_{03}(s)S_{12}(u)\,du\,ds + \int_{t_{0M}}^{t_{4m}} f_{03}(s)(1-S_{34}(t_{4m}-s))S_{01}(s)\,ds\}+$$

$$\sum_{k=0}^{n_{014}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{4m}-s} \int_{0}^{t_{4m}-s-u} f_{01}(s)f_{12}(u)f_{23}(r)(1-S_{34}(t_{4m}-s-u-r))S_{03}(s)S_{13}(u)S_{24}(r)\,dr\,du\,ds+$$
$$\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{4m}-s} \int_{0}^{t_{4m}-s-u} f_{01}(s)f_{12}(u)f_{24}(r)S_{03}(s)S_{13}(u)S_{23}(r)\,dr\,du\,ds+$$
$$\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{4m}-s} f_{01}(s)f_{13}(u)(1-S_{34}(t_{4m}-s-u))S_{03}(s)S_{12}(u)\,du\,ds\}+$$
$$\sum_{k=0}^{n_{024}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} \int_{t_{2M}-s-u}^{t_{4m}-s-u} f_{01}(s)f_{12}(u)f_{23}(r)(1-S_{34}(t_{4m}-s-u-r))S_{03}(s)S_{13}(u)S_{24}(r)\,dr\,du\,ds+$$
$$\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} \int_{t_{2M}-s-u}^{t_{4m}-s-u} f_{01}(s)f_{12}(u)f_{24}(r)S_{03}(s)S_{13}(u)S_{23}(r)\,dr\,du\,ds\}+$$
$$\sum_{k=0}^{n_{034}} \log\{\int_{t_{0M}}^{t_{3m}} \int_{0}^{t_{3m}-s} \int_{0}^{t_{3m}-s-u} f_{01}(s)f_{12}(u)f_{23}(r)(S_{34}(t_{3M}-s-u-r)-S_{34}(t_{4m}-s-u-r))S_{03}(s)S_{13}(u)S_{24}(r)\,dr\,du\,ds+$$
$$\int_{t_{0M}}^{t_{3m}} \int_{0}^{t_{3m}-s} f_{01}(s)f_{13}(u)(S_{34}(t_{3M}-s-u)-S_{34}(t_{4m}-s-u))S_{03}(s)S_{12}(u)\,du\,ds+$$
$$\int_{t_{0M}}^{t_{3m}} f_{03}(s)(S_{34}(t_{3M}-s)-S_{34}(t_{4m}-s))S_{01}(s)\,ds\}+$$
$$\sum_{k=0}^{n_{0124}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} \int_{t_{2M}-s-u}^{t_{4m}-s-u} f_{01}(s)f_{12}(u)f_{23}(r)(1-S_{34}(t_{4m}-s-u-r))S_{03}(s)S_{13}(u)S_{24}(r)\,dr\,du\,ds+$$
$$\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{2m}-s} \int_{t_{2M}-s-u}^{t_{4m}-s-u} f_{01}(s)f_{12}(u)f_{24}(r)S_{03}(s)S_{13}(u)S_{23}(r)\,dr\,du\,ds\}+$$
$$\sum_{k=0}^{n_{0134}} \log\{\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{3m}-s} \int_{0}^{t_{3m}-s-u} f_{01}(s)f_{12}(u)f_{23}(r)(S_{34}(t_{3M}-s-u-r)-S_{34}(t_{4m}-s-u-r))S_{03}(s)S_{13}(u)S_{24}(r)\,dr\,du\,ds+$$
$$\int_{t_{0M}}^{t_{1m}} \int_{t_{1M}-s}^{t_{3m}-s} f_{01}(s)f_{13}(u)(S_{34}(t_{3M}-s-u)-S_{34}(t_{4m}-s-u))S_{03}(s)S_{12}(u)\,du\,ds\}+$$
$$\sum_{k=0}^{n_{0234}} \log\{\int_{t_{0M}}^{t_{2m}} \int_{0}^{t_{2m}-s} \int_{t_{2M}-s-u}^{t_{3m}-s-u} f_{01}(s)f_{12}(u)f_{23}(r)(S_{34}(t_{3M}-s-u-r)-S_{34}(t_{4m}-s-u-r))S_{03}(s)S_{13}(u)S_{24}(r)\,dr\,du\,ds\}.$$