JIRSS (2023)
Vol. 22, No. 02, pp 35-58
DOI: 10.22034/jirss.2024.1999145.1011

Some Applications of Phase-type Distributions in Recurrent Events

Raoufeh Asghari [1], **Amin Hassan Zadeh** [2]

[1] Department of Statistics, Shahid Beheshti University, Tehran, Iran
[2] Department of Mathematics, Oregon State University, Corvallis, OR, USA.

Received: 28/03/2023, Accepted: 25/05/2024, Published online: 30/10/2024

Abstract. In this paper, the recurrent events that can occur more than once over the follow-up time have been modeled by phase-type distributions. The time till death is assumed to have a phase-type distribution (defined in a Markov chain environment) with interpretable parameters. We use phase-type distribution to calculate the probability of various number of transitions, the conditional expected time to stay in a disease stage and the probability of transition from a stage to another. Using both real and simulated datasets, the model has been calibrated. The confidence intervals for the parameters have been created using bootstrap approaches.

Keywords. Multiple State Models, Recurrent Events, Phase-type Distribution, Bootstrap, Cancer, Stanford Heart Transplant Data.

MSC: 92B15, 60J22.

1 Introduction

Multi-state models are the most often used models for describing longitudinal survival data. According to Andersen et al. (1993), a multi-state model is a stochastic process model with a number of states and the potential for transitions between them. The states during the duration of a follow-up show the person's varied situations (healthy, ill, etc.). Some specific multi-state models that have been widely used in biological applications include the illness-death model, the bivariate model, and the three-state progressive model Hougaard (2000). Reversible illness-death Moreira (2013) is a

Raoufeh Asghari (Raoufe.Asghari@gmail.com)
Corresponding Author : Amin Hassan Zadeh (hassanza@oregonstate.edu)

model that is used in this study and is one of the key characteristics of many state models.

The event of interest may repeat again during the patient follow-up period in many biological investigations. These occurrences are referred to as recurrent events Kleinbaum and Klein (2012). Recurrent occurrences are those that happen repeatedly and are of the same sort. Examples include acute asthmatic flare-ups in young children, epileptic seizures, cancer recurrences, myocardial infarctions, migraine discomfort, and ear infections. In the literature, several statistical models, including Andersen and Gill (1982), Lin et al. (2000), Kelly and Lim (2000), and Cook and Lawless (2010), have been presented to study recurrent events. The R package TPmsm, which seeks to construct nonparametric and semiparametric estimators for the transition probabilities in progressive illness-death models, is described in Araujo et al. (2014).

Another R package called TP.idm, which implements a novel non-Markovian estimator for the transition probability matrix in the progressive illness-death model under right censoring, is also described in Balboa and Una-Alvarez (2018). Various modeling techniques are reviewed in Meira-Machado et al. (2009) that follow the multi-state model methodology with an emphasis on the estimate of various parameters like transition probabilities and survival probabilities.

In this study, we employ a continuous time Markov process with a finite-state and a single absorbing state. The state space in the Markov process is partitioned into some subspaces. Each subspace will represent a different stage in a person's health. Each subspace, for instance, is read as either a stage of the malignancy or a stage of recovery in the case of cancer. Each subspace will have states that allow the models to reflect the semi-Markov attribute. This method sets our model apart from earlier Markovian models applied to recurrent events. In Asghari and Hassan Zadeh (2020) this technique has been applied in a different setting.

As the time to death follows a phase-type (PH) distribution, we can take advantages of the PH properties. In the class of all distributions defined on the non-negative real numbers, the set of PH is dense. It can be used to approximate any positive-valued distribution, see Asmussen et al. (1996). There are closed-form expressions for the distribution and the probability density functions, and this also applies to the Laplace–Stieltjes transform and we can get mean and the all non-central moments by successive differentiation of the Laplace transform. Since the PH is defined based on an underlying Markov process, its parameters (which are called the representation) are the initial probability vector and the sub-intensity matrix of the underlying Markov chain. Therefore, the representation of a PH might be interpretable. We have used this advantage of the PH in our model.

Recently, a lot of researchers have employed the PH distribution. The PH is used by Hassan Zadeh et al. (2013) to model disability insurance. According to the paradigm, aging is the progression through many stages of declining vitality. When disabled, people also go through a number of stages that signify the length of their disability.

The *PH* is used by Asghari and Hassan Zadeh (2020) to model skin cancer patients in the United States to estimate factors linked to aging that may be compared with the physiological aging processes of cancer patients and healthy individuals. Phase-type distribution is used for mortality analysis in Lin and Liu (2007), and are obtained conditional survival probabilities of the time of death and the actuarial present values of the whole life insurance. Phase-type distribution is used for analyzing data on lengths of stay of hospital patients in Faddy and McClean (2000) and also Aalen (1995) uses *PH* to model the interval time between first and second birth. A special case of *PH* is Coxian distribution that can be used to represent survival times in terms of phases through which an individual may progress until leaves a system, such as hospital stay or time till death, Marshall and Zenga (2009) used the Coxian to model the patients stay in hospital.

This paper is organized as follows: Section 2 provides an introduction to *PH* as well as the notational convention that will be used in describing our model. In Section 3, we describe the structure of our model and present the main theorem which is the main contribution of this paper. In this section two examples will be presented. Section 4 closes the paper with some concluding remarks and a brief discussion of possible future studies.

2 Definition of Phase-type Distributions

Consider a continuous-time Markov process $\{J_t, t \geq 0\}$ with the space state, $\Gamma = \{0, 1, \ldots, m\}, m \in \mathbb{N}$, where the state 0 is absorbing and the rest are transient. Assume the intensity matrix and initial probability vector associated with J, are denoted by $\mathbf{Q}$ and $\boldsymbol{\beta}$, respectively. Clearly, we have that

$$\mathbf{Q} = \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{t}_0 & \mathbf{T} \end{bmatrix}, \tag{1}$$

where the matrix $\mathbf{T}$ is $m \times m$, sub–intensity matrix (matrix $\mathbf{B} = (b_{ij})_{i,j=1,\ldots,m}$ is called a sub–intensity matrix if $b_{ii} \leq 0$, $b_{ij} \geq 0$, and $\sum_{k=1}^{m} b_{ik} \leq 0$, with strict inequality for at least on i, for $i, j = 1, 2, \ldots, m;\ i \neq j$). The elements of the m- dimensional column vector $\mathbf{t}_0$ are the transition rates from the transient states to the absorbing state, 0. Therefore we have $\mathbf{T1} + \mathbf{t}_0 = \mathbf{0}$, (where $\mathbf{0}$ and $\mathbf{1}$ are column vectors of zeros and ones with proper dimensions, respectively). It can be easily proved that

$$e^{\mathbf{Q}y} = \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{1} - e^{\mathbf{T}y}\mathbf{1} & e^{\mathbf{T}y} \end{bmatrix}, \tag{2}$$

(matrix exponential of the squared matrix $\mathbf{B}$ is defined as $e^{\mathbf{B}} = \sum_{l=0}^{\infty} \frac{\mathbf{B}^l}{l!}$), for a proof, see Latouche and Ramaswami (1999).

We also assume that $P(J_0 = 0) = \beta_0 = 0$. As a result, $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\alpha}')$ where $\boldsymbol{\alpha}$ is a column vector of size m and its transpose is denoted by $\boldsymbol{\alpha}'$, so we have $\boldsymbol{\alpha}'\mathbf{1} = 1$. In this paper,

all vectors are columns and the transpose of vector $\boldsymbol{\alpha}$ is $\boldsymbol{\alpha}'$

If we define $Y = \inf\{t : J_t = 0\}$, then it is said that Y is a *PH* random variable with representation $(\boldsymbol{\alpha}, \mathbf{T})$. For a continuous random variable Y, the event $Y > y$ implies that the process started from any transient state has not reached the absorbing state by time y, thus the distribution function of Y can be obtained as follows:

$$F(y) = 1 - \boldsymbol{\alpha}' e^{\mathbf{T}y}\mathbf{1}, \quad y \geq 0.$$

Thus the survival function of Y is

$$S(y) = \boldsymbol{\alpha}' e^{\mathbf{T}y}\mathbf{1}, \quad y \geq 0. \tag{3}$$

By taking derivatives of $F(y)$, one can obtain the probability density function of Y as follows,

$$f_Y(y) = \boldsymbol{\alpha}' e^{\mathbf{T}y}\mathbf{t}_0, \quad y \geq 0.$$

and Laplace-transform and k^{th} moments are given in the following.

$$\Phi(s) = \boldsymbol{\alpha}'(s\mathbf{I} - \mathbf{T})^{-1}\mathbf{t}_0,$$
$$E(Y^k) = k!\boldsymbol{\alpha}'(-\mathbf{T}^{-1})^k\mathbf{1}, \quad k = 0, 1, 2, \cdots.$$

For a complete review of *PH* distributions refer to Neuts (1994).

3 Transition Probabilities

Based on the structure of a representation of the *PH* several classes of phase-type distributions can be distinguished. The structure of a *PH* representation often has an impact on its application, as some structures allow more efficient solutions. The most important distinction is the one into acyclic and General Phase-type distributions: at least one Markovian representation without cycles in the sub-generator exists for every acyclic *PH*, whereas cycles are permitted for generic phase-type distributions, see Reinecke et al. (2012). In the reversible illness-death model, the general PH is applied. The reversible illness-death model in a Markovian setting is examined in this study.

The analysis in recurrent studies is often performed using the multi–state models. These models are very useful for describing event history data offering a better understanding of the process of the illness, and leading to a better knowledge of the evolution of the disease over time. The complexity of a multi-state model greatly depends on the number of states defined and the transitions allowed between these states. Our approach is to calculate the probability of the number of transitions.

We use the finite-state continuous-time Markov process with multi states. One absorbing state is the death phase and the other states are the recovery and disease phases. From now on, we will use the word "stage" rather than "state" or "phase" in

our multi-state model. In fact, we will assume k stages (from 1 to k) for the disease and recovery, (one stage for the recovery and $k-1$ for the disease) and one absorbing state for death. Each stage will have n transient states. In this model transition from every stage to another is possible. In some cases, the states inside stages can be interpreted as physiological ages, see Lin and Liu (2007) and Asghari and Hassan Zadeh (2020) where aging is transition from one physiological age to the next physiological age and process will end when transition occurs from any other state to the absorbing state, death. For each state $i, i = 1, 2, .., n$ in every stage, several parameters are used for modeling mortality. Issues of interest *i.e.*, the number of recurrences or transitions until time t, the number of transitions from one stage to another and the expected time stay in every stage will be calculated.

After the diagnosis of a special disease, that stage is also known. So patient is at one of the disease stages at time $t = 0$ -arrival time- at age x. At any time thereafter, he may arrive at a death state, recover or visit another disease stage. Recovered patient may become sick or dead. Our proposed multi-state model is reversible, in the sense that past stages can be revisited. For this model, let E denote the space state of the underlying Markov chain, then $E = \bigcup_{j=1}^{k} E_i \bigcup D$ where sets of $E_1, E_2, \ldots, E_k$ are to represent stages, including recovery, and the state $D = \{0\}$ represents death. Every stage, including recovery, has a finite set of n states labeled $1, ..., n$, with instantaneous transitions being possible between selected pairs of states, see Figure 1. In general, there are $k \times n + 1$ states, where $k \times n$ are for recovery and disease stages and 1 for the death state.

For each $t \geq 0$, the continuous Markovian random process J_t is in one of the stages $1, ..., k$ or in D. We interpret the event $J_t \in E_i$ to mean that the individual is in stage i at age $x + t$, $i = 1, \ldots, k$. We assume the rate of transition from the disease stages depends on stage, age and duration of illness.

As mentioned earlier, we propose a reversible illness–death model involving n states in each alive stage. Patients may move to the next stage of disease, recovery or death. The recovery rate will normally be lower for the later stages of the disease. Of course, patients may die while in any state of stages, as indicated by the arrows to death in Figure 1.

Since the Markov process has only a single absorbing state, the time of death (the time of absorbing) follows a *PH* distribution. This means that the patient before his death is either in a stage of recovery or disease, so it can be said that the time of staying in these situations can be calculated via the Markov property.

Furthermore, the sub-intensity matrix, $\mathbf{T}$, of the Markov process in (1) is given by

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_{1,2} & \mathbf{T}_{1,3} & \cdots & \mathbf{T}_{1,k} \\ \mathbf{T}_{2,1} & \mathbf{T}_2 & \mathbf{T}_{2,3} & \cdots & \mathbf{T}_{2,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{T}_{k,1} & \mathbf{T}_{k,2} & \mathbf{T}_{k,3} & \cdots & \mathbf{T}_k \end{bmatrix}, \tag{4}$$

where $\mathbf{T}_i, i \neq j,\ i,j = 1,2,...,k$ are $n \times n$ matrices containing the transition rates from stage i to j. The matrices $\mathbf{T}_i$'s, $i = 1,2,...,k$ are the sub-intensity matrices (with non-zero main and upper diagonal elements) for the Markov chain describing a sojourn in stage i. The mortality rates, $\mathbf{t}_0 = -\mathbf{T1}$, is an $n \times k$-dimensional column vector containing the rate of death from each of the $n \times k$ alive states and $\mathbf{1}$ is a column vector of 1's with proper dimension.

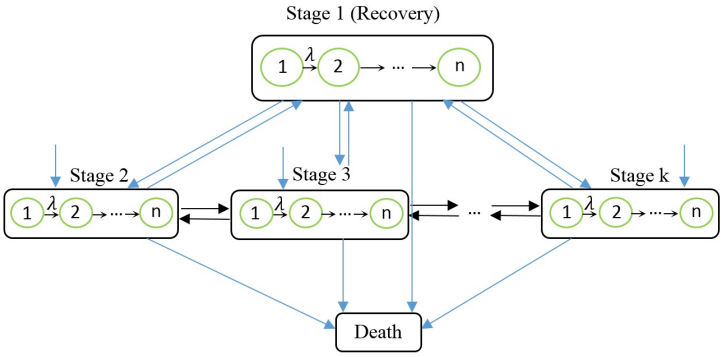

Figure 1: Reversible illness-death model: the k+1 stages (boxes) and the possible transition among them (arrows).

In this paper, we assume that the person at diagnosis age x and we will remove x from the notations.

From now on, it is also assumed that $\boldsymbol{\alpha}' = (\boldsymbol{\alpha_1}', \boldsymbol{\alpha_2}', \ldots, \boldsymbol{\alpha_k}')$, where $\boldsymbol{\alpha_i}$'s are all $n \times 1$ vectors corresponding to initial probabilities to start from stage i ($\boldsymbol{\alpha_1}$ corresponds to the recovery which can be assumed to be $\mathbf{0}$). The normalized vector of vector $\boldsymbol{\beta}$ is denoted by $\hat{\boldsymbol{\beta}}$.

Now with the notations above and assumptions, by using the Markov property we are able to calculate the probability of various number of transitions until time t.

Theorem 3.1. *Let $\{J_t, t \geq 0\}$ be a continuous-time Markov chain with space state $E = \bigcup_{j=1}^{k} E_j \cup D$ and with the sub-intensity matrix* (4) *where D the only absorbing state and E_j's are mutual disjoint subsets of E. Assume $\{N(t), t \geq 0\}$ is the number of transitions between E_j's in $[0,t]$ or from any E_j to D, $j = 1,...,k$, and define $P_t^{(i)}(l) = P[N(t) = l \mid J_0 \in E_i]$. Then we have the following:*

$$P_t^{(i)}(0) = \hat{\boldsymbol{\alpha}}_i' \, e^{\mathbf{T}_i t}\mathbf{1}, \tag{5}$$

$$P_t^{(i)}(1) = \sum_{\substack{i_1=1 \\ i_1 \neq i}}^{k} \hat{\boldsymbol{\alpha}}_i' \, \mathbf{x}_{\{i_1\}}^{(i)}(t)\mathbf{1}, \tag{6}$$

$$P_t^{(i)}(2) = \sum_{\substack{i_1,i_2 \\ i_1 \neq i \\ i_2 \neq i_1}}^{k} \hat{\boldsymbol{\alpha}}_i' \mathbf{x}_{\{i_1,i_2\}}^{(i)}(t)\mathbf{1}, \tag{7}$$

$$\vdots$$

$$P_t^{(i)}(l) = \sum_{\substack{i_1,i_2,\ldots,i_l \\ i_1 \neq i \\ i_2 \neq i_1 \\ \vdots \\ i_l \neq i_{l-1}}}^{l} \hat{\boldsymbol{\alpha}}_i' \mathbf{x}_{\{i_1,i_2\cdots,i_l\}}^{(i)}(t)\mathbf{1}, \tag{8}$$

where $\mathbf{x}_{\{i_1\}}^{(i)}(.), \mathbf{x}_{\{i_1 i_2\}}^{(i)}(.), \cdots, \mathbf{x}_{\{i_1 i_2 \cdots i_k\}}^{(i)}(.)$ *satisfy the following differential equations:*

$$\frac{d}{dt}\mathbf{x}_{\{i_1\}}^{(i)}(t) = e^{\boldsymbol{T}_i t}\boldsymbol{T}_{i,i_1} + \mathbf{x}_{\{i_1\}}^{(i)}(t)\boldsymbol{T}_{i_1},$$

$$\frac{d}{dt}\mathbf{x}_{\{i_1,i_2\}}^{(i)}(t) = \mathbf{x}_{\{i_1\}}^{(i)}(t)\boldsymbol{T}_{i_1,i_2} + \mathbf{x}_{\{i_1,i_2\}}^{(i)}(t)\boldsymbol{T}_{i_2},$$

and

$$\frac{d}{dt}\mathbf{x}_{\{i_1,i_2,\cdots,i_l\}}^{(i)}(t) = \mathbf{x}_{\{i_1,i_2\cdots i_{l-1}\}}^{(i)}(t)\boldsymbol{T}_{i_{l-1},i_l} + \mathbf{x}_{\{i_1,i_2,\cdots,i_l\}}^{(i)}(t)\boldsymbol{T}_{i_l},$$

with the matrices $\mathbf{x}$ *satisfy the following recursive equations*

$$\mathbf{x}_{\{i_1,i_2,\ldots,i_{m-1},i_m\}}^{(i)}(t) = \int_0^t \mathbf{x}_{\{i_1,i_2,\ldots,i_{m-1}\}}^{(i)}(z)\boldsymbol{T}_{i_{m-1}i_m} e^{\boldsymbol{T}_{i_m}(t-z)}dz,$$

for $m = 1, 2, \ldots,$ *where*

$$\mathbf{x}_{\{\}}^{(i)}(t) = e^{\boldsymbol{T}_i t},$$

and all $\mathbf{x}$*'s are zero matrices, with proper dimension, at* $t = 0$*. Notation* $\{\}$ *is used to denote the null set.*

Proof: The Equation (5) is obvious. To prove the rest, we divide $[0, t]$ into m equal subintervals, each length will be $ds = \frac{t}{m}$. First, we assume that any transition between the stages will occur in a multiplier of ds. Assume also that $s_\kappa = \kappa\, ds$, $\kappa = 0, 1, \ldots,$ then

we have the followings:

$$\begin{aligned}
P_t^{(i)}(1) &= Pr[N(t)=1 \mid J_0 \in E_i] = E[\mathbf{1}_{N(t)=1} \mid J_0 \in E_i] \\
&= E[\mathbf{1}_{\{\bigcup_{\substack{i_1=1\\ i_1\neq i}}^{k} \bigcup_{\kappa=0}^{m-1} \underset{\forall\, u\in[0,s_\kappa]}{J_u\in E_i}\,,\, J_{s_\kappa+ds}\in E_{i_1},\ \underset{\forall\, v\in[0,t-(s_\kappa+ds)]}{J_{s_\kappa+ds+v}\in E_{i_1}}\}} \mid J_0 \in E_i] \\
&= E[\sum_{\substack{i_1=1\\ i_1\neq i}}^{k}\sum_{\kappa=0}^{m-1} \mathbf{1}_{\{\underset{\forall\, u\in[0,s_\kappa]}{J_u\in E_i}\,,J_{s_\kappa+ds}\in E_{i_1},\ \underset{\forall\, v\in[0,t-(s_\kappa+ds)]}{J_{s_\kappa+ds+v}\in E_{i_1}}\}} \mid J_0 \in E_i] \\
&= \sum_{\substack{i_1=1\\ i_1\neq i}}^{k}\sum_{\kappa=0}^{m-1} E[\mathbf{1}_{\{\underset{\forall\, u\in[0,s_\kappa]}{J_u\in E_i}\,,J_{s_\kappa+ds}\in E_{i_1},\ \underset{\forall\, v\in[0,t-(s_\kappa+ds)]}{J_{s_\kappa+ds+v}\in E_{i_1}}\}} \mid J_0 \in E_i] \\
&= \sum_{\substack{i_1=1\\ i_1\neq i}}^{k}\sum_{\kappa=0}^{m-1} Pr[\underset{\forall\, u\in[0,s_\kappa]}{J_u \in E_i}\,, J_{s_\kappa+ds}\in E_{i_1},\ \underset{\forall\, v\in[0,t-(s_\kappa+ds)]}{J_{s_\kappa+ds+v}\in E_{i_1}} \mid J_0 \in E_i] \\
&= \sum_{\substack{i_1=1\\ i_1\neq i}}^{k}\sum_{\kappa=0}^{m-1} Pr[\underset{\forall\, u\in[0,s_\kappa]}{J_u \in E_i} \mid J_0\in E_i]Pr[J_{s_\kappa+ds}\in E_{i_1}\mid J_{s_\kappa}\in E_i]Pr[\underset{\forall\, v\in[0,t-(s_\kappa+ds)]}{J_{s_\kappa+ds+v}\in E_{i_1}} \mid J_{s_\kappa+ds}\in E_{i_1}].
\end{aligned}$$

For $i, i_1 = 1, ..., k, D$, $i \neq i_1$.

Now, by using equation (5), and letting $ds \to 0$ we end up with the followings:

$$P_t^{(i)}(1) = \sum_{\substack{i_1=1\\ i_1\neq i}}^{k} \lim_{ds\to 0} \sum_{\kappa=0}^{m-1} (\hat{\boldsymbol{\alpha}}_i' e^{\mathbf{T}_i \kappa\, ds} \mathbf{I}'_{E,E_i})(e^{\mathbf{T} ds}\mathbf{I}_{E,E_{i_1}})(e^{\mathbf{T}_{i_1}(t-(\kappa+1)ds)}\mathbf{1}), \tag{9}$$

from $\lim_{\epsilon\to 0^+} \frac{e^{\mathbf{T}\epsilon}-\mathbf{I}}{\epsilon} = \mathbf{T}$, the continuity of $e^{\mathbf{T}s}$ and the limit of the Riemann sum, (9) becomes

$$\begin{aligned}
P_t^{(i)}(1) &= \sum_{\substack{i_1=1\\ i_1\neq i}}^{k} \int_0^t \hat{\boldsymbol{\alpha}}_i' e^{\mathbf{T}_i s}\mathbf{T}_{i,i_1} e^{\mathbf{T}_{i_1}(t-s)}\mathbf{1} ds \\
&= \sum_{\substack{i_1=1\\ i_1\neq i}}^{k} \hat{\boldsymbol{\alpha}}_i' \int_0^t e^{\mathbf{T}_i s}\mathbf{T}_{i,i_1} e^{\mathbf{T}_{i_1}(t-s)} ds \mathbf{1} \\
&= \sum_{\substack{i_1=1\\ i_1\neq i}}^{k} \hat{\boldsymbol{\alpha}}_i' \mathbf{x}^{(i)}_{\{i_1\}}(t)\mathbf{1}.
\end{aligned} \tag{10}$$

Where $\mathbf{x}^{(i)}_{\{i_1\}}(t) = \int_0^t e^{\mathbf{T}_i s}\mathbf{T}_{i,i_1} e^{\mathbf{T}_{i_1}(t-s)} ds$. It can be seen that

$$\frac{d}{dt}\mathbf{x}^{(i)}_{\{i_1\}}(t) = e^{\mathbf{T}_i t}\mathbf{T}_{i,i_1} + \mathbf{x}^{(i)}_{\{i_1\}}(t)\mathbf{T}_{i_1}.$$

Note that $\frac{de^{\mathbf{T}z}}{dz} = \mathbf{T}e^{\mathbf{T}}$. The event of $N(t) = 2$ means there is 2 transitions during interval $[0, t]$. Assuming the transitions happen at time s_{κ_1} and s_{κ_2}, we have the following:

$$
\begin{aligned}
P_t^{(i)}(2) &= Pr\left(N(t) = 2 \mid J_0 \in E_i\right) = E\left(\mathbf{1}_{N(t)=2} \mid J_0 \in E_i\right) \\
&= E\left(\mathbf{1}_{\bigcup_{\substack{i_1,i_2 \\ i_1\neq i_2}}^{k} \bigcup_{\kappa_1=0}^{m-2} \bigcup_{\kappa_2=\kappa_1+1}^{m-1} \underset{\forall\, u\in[0,s_{\kappa_1}]}{J_u\in E_i}\ J_{s_{\kappa_1}+ds}\in E_{i_1},\ \underset{\forall\, v\in[0,s_{\kappa_2}-(s_{\kappa_1}+ds)]}{J_{s_{\kappa_1}+ds+v}\in E_{i_1}},\ J_{s_{\kappa_2}+ds}\in E_{i_2},\ \underset{\forall\, \omega\in[0,t-(s_{\kappa_2}+ds)]}{J_{s_{\kappa_2}+ds+\omega}\in E_{i_2}}}\right) \\
&= \sum_{\substack{i_1,i_2 \\ i_1\neq i \\ i_2\neq i_1}}^{k} \sum_{\kappa_1=0}^{m-2} \sum_{\kappa_2=\kappa_1+1}^{m-1} Pr[\underset{\forall\, u\in[0,s_{\kappa_1}]}{J_u \in E_i} \mid J_0 \in E_i]\, Pr[J_{s_{\kappa_1}+ds} \in E_{i_1} \mid J_{s_{\kappa_1}} \in E_i] \\
&\quad Pr[\underset{\forall\, v\in[0,s_{\kappa_2}-(s_{\kappa_1}+ds)]}{J_{s_{\kappa_1}+ds+v} \in E_{i_1}} \mid J_{s_{\kappa_1}+ds} \in E_{i_1}]\, Pr[J_{s_{\kappa_2}+ds} \in E_{i_2} \mid J_{s_{\kappa_2}} \in E_{i_1}] \\
&\quad Pr[\underset{\forall\, \omega\in[0,t-(s_{\kappa_2}+ds)]}{J_{s_{\kappa_2}+ds+\omega} \in E_{i_2}} \mid J_{s_{\kappa_2}+ds} \in E_{i_2}].
\end{aligned}
$$

As $ds \to 0$ and using the same techniques used in (10) we will end with the following.

$$
\begin{aligned}
P_t^{(i)}(2) &= \sum_{\substack{i_1,i_2 \\ i_1\neq i \\ i_2\neq i_1}}^{k} \int_0^t \int_0^z \hat{\boldsymbol{\alpha}}_i' e^{\mathbf{T}_i s} \mathbf{T}_{i\,i_1} e^{\mathbf{T}_{i_1}(z-s)} \mathbf{T}_{i_1 i_2} e^{\mathbf{T}_{i_2}(t-z)} \mathbf{1}\, ds dz \\
&= \sum_{\substack{i_1,i_2 \\ i_1\neq i \\ i_2\neq i_1}}^{k} \hat{\boldsymbol{\alpha}}_i' \int_0^t \int_0^z e^{\mathbf{T}_i s} \mathbf{T}_{i\,i_1} e^{\mathbf{T}_{i_1}(z-s)} \mathbf{T}_{i_1 i_2} e^{\mathbf{T}_{i_2}(t-z)} \mathbf{1}\, ds dz \\
&= \sum_{\substack{i_1,i_2 \\ i_1\neq i \\ i_2\neq i_1}}^{k} \hat{\boldsymbol{\alpha}}_i' \int_0^t \mathbf{x}_{\{i_1\}}^{(i)}(z) \mathbf{T}_{i_1 i_2} e^{\mathbf{T}_{i_2}(t-z)} \mathbf{1} dz \\
&= \sum_{\substack{i_1,i_2 \\ i_1\neq i \\ i_2\neq i_1}}^{k} \hat{\boldsymbol{\alpha}}_i' \mathbf{x}_{\{i_1\,i_2\}}^{(i)}(t) \mathbf{1}.
\end{aligned}
$$

Where $\mathbf{x}_{\{i_1\,i_2\}}^{(i)}(t)$ satisfies the following differential equation

$$
\frac{d\mathbf{x}_{\{i_1\,i_2\}}^{(i)}(t)}{dt} = \mathbf{x}_{\{i_1\}}^{(i)}(t) \mathbf{T}_{i_1\,i_2} + \mathbf{x}_{\{i_1\,i_2\}}^{(i)}(t) \mathbf{T}_{i_2}.
$$

In general, it can be seen easily that $P_t^{(i)}(l)$ equals

$$P_t^{(i)}(l) = \sum_{\substack{i_1,i_2,\ldots,i_l \\ i_1\neq i \\ i_2\neq i_1 \\ \vdots \\ i_l\neq i_{l-1}}}^{k} \int_0^t \int_0^{z_{l-1}} \cdots \int_0^{z_1} \hat{\boldsymbol{\alpha}}_i' e^{\mathbf{T}_i s} \mathbf{T}_{i i_1} e^{\mathbf{T}_{i_1}(z_1-s)} \mathbf{T}_{i_1 i_2} e^{\mathbf{T}_{i_2}(z_2-z_1)} \ldots \mathbf{T}_{i_{l-1} i_l} e^{\mathbf{T}_{i_l}(t-z_{l-1})} \mathbf{1}\, ds\, dz_1\, dz_2 \cdots dz_{l-1}$$

$$= \sum_{\substack{i_1,i_2 \\ i_1\neq i \\ i_2\neq i_1}}^{k} \hat{\boldsymbol{\alpha}}_i' \int_0^t \mathbf{x}^{(i)}_{\{i_1,i_2,\cdots,i_{l-1}\}}(z_{l-1}) \mathbf{T}_{i_{l-1} i_l} e^{\mathbf{T}_{i_l}(t-z_{l-1})} \mathbf{1} dz_{l-1}$$

$$= \sum_{\substack{i_1,i_2 \\ i_1\neq i \\ i_2\neq i_1}}^{k} \hat{\boldsymbol{\alpha}}_i' \mathbf{x}^{(i)}_{\{i_1,i_2,\cdots,i_l\}}(t)\mathbf{1},$$

where $\mathbf{x}^{(i)}_{\{i_1,i_2,\cdots,i_l\}}(t)$ satisfies the following differential equation

$$\frac{d\mathbf{x}^{(i)}_{\{i_1,i_2,\cdots,i_l\}}(t)}{dt} = \mathbf{x}^{(i)}_{\{i_1,i_2\cdots i_{l-1}\}}(t)\mathbf{T}_{i_{l-1},i_l} + \mathbf{x}^{(i)}_{\{i_1,i_2,\cdots,i_l\}}(t)\mathbf{T}_{i_l}.$$

To solve the differential equations above, we use ODE-45 function in MATLAB (2018) software. In the following we will consider two examples for the applications.

Example 1: The Stanford Heart Transplantation Study

In October 1967, the Stanford heart transplant project got underway. This data collection is referenced in either Crowley and Hu (1977) or Kalbfleisch and Prentice (1980) (pages 230–232). The available data covers the period until April 1, 1974. Some patients passed away before the right heart was found. Of the 103, 69 had heart transplants, and 45 (or 65%) of them died as a result. Thirty (88%) of the 34 patients without transplants passed away.

The 28 patients who were still alive provided censored survival times. A vector of covariates including age at acceptance (Age), year of acceptance (Year), previous surgery record (Surgery: coded as 1 = yes; 0 = no), and transplant (Transplant: coded as 1 = yes; 0 = no) were recorded for each individual. Additionally, an indicator of each person's final vital status (censored or not) and the survival times (time to transplant, time to death) from the patient's entry into the study (in days) are included. The sole time-dependent covariate is the Transplant.

Three variables—start, stop, and event—are used in this data structure to convey an individual's survival. The time-dependent covariate "transplant" in the Stanford study

stands in for a treatment intervention. Patients with a change in the time-dependent covariate must be represented by two lines of data in contrast to individuals who do not experience a change in the time-dependent covariate. The first line for these patients denotes the time until the transplant, and the second line denotes the time from the transplant until the end of the follow-up or death. The remaining (time-fixed) covariates are the same for the two lines. The variables start and stop identify the start and end of the data's time interval for each row, whereas event is an indication variable with a value of 1 if a death occurred at time stop and 0 otherwise.

For clarification, we repeat the example in Meira-Machado et al. (2009). Consider the information available from four patients (from the Stanford study) with identification numbers 25, 26, 27 and 28 in Table 1. For the first two patients the time from enrollment to censoring is 1800 and 1401 days, respectively, and the first patient had a heart transplant 25 days after enrollment. The time from enrollment to death for the third and fourth patients are 263 and 72 days, respectively, and the last patient received a new heart on day 71.

Table 1: Stanford heart transplantation in Example 2. Meira-Machado et al. (2009)

id	start	stop	event	transplant	age	year	surgery
25	0	25	0	0	33.2238	1.57426	0
25	25	1800	0	1	33.2238	1.57426	0
26	0	1401	0	0	30.5353	1.58248	0
27	0	263	1	0	8.7858	1.59069	0
28	0	71	0	0	54.0233	1.68378	0
28	71	72	1	1	54.0233	1.68378	0

The descriptive statistics of the data by Age (intervals of 5-year), Year (from 1967 to 1974, intervals of 1) and Surgery (yes=1 or no=0) are presented in Tables 2, 3 and 4, respectively.

Table 2: The descriptive statistics of the data by age in Example 2.

Age	N	Trans.	Death	$\text{Death}_{\text{trans.}}$	$\%\ \text{Death}_{\text{without trans.}}$	$\%\ \text{Death}_{\text{after trans.}}$
<25	5	2	4	1	100	50
25-30	5	4	1	1	0	25
30-35	4	2	1	0	50	0
35-40	7	4	4	1	100	25
40-45	19	12	15	9	100	75
45-50	29	22	19	13	86	59
50-55	27	17	23	14	90	82
>55	8	6	8	6	100	100
sum	103	69	75	45		

In Table 2, the average age of heart patients is 45.2 years. The percent of trans-

planted patients 45-50 years old is the maximum with 78% transplanting. The rate of death for transplanted patients ($\frac{\#\text{death of transplanted patients}}{\#\text{ transplanted patients}}$) increases with age.

Table 3: The descriptive statistics of the data by year in Example 2.

Year	N	Trans.	Death	$\text{Death}_{trans.}$	% $\text{Death}_{\text{without trans.}}$	% $\text{Death}_{\text{after trans.}}$
1	16	7	16	7	100	100
2	17	11	15	10	83	91
3	10	8	6	4	100	50
4	17	11	14	8	100	73
5	17	12	12	8	100	67
6	19	16	11	8	100	50
>6	7	4	1	0	33	0
sum	103	69	75	45		

In Table 3, the number of heart patients' enrollment is about the same for every year but the percent of transplanted patients for the fifth year of study is maximum with 84% transplanting. The mortality rate for transplanted patients has decreased over the years so it is expected that the influence of year of acceptance on hazard is negative.

Table 4: The descriptive statistics of the data by surgery in Example 2.

surgery	N	Trans.	Death	$\text{Death}_{\text{trans.}}$	% $\text{Death}_{\text{without trans.}}$	% $\text{Death}_{\text{after trans.}}$
0	87	56	66	39	87	70
1	16	13	9	6	100	46
sum	103	69	75	45		

As can be seen in Table 4, most patients have no history of surgery and patients without surgery are more likely to die, about 76%. Heart patients with surgery are more likely to have heart transplants than those without heart surgery, in other words, 81% of patients with surgery and 64% of patients without surgery have a heart transplant. Therefore, it is expected that the influence of previous surgery on hazard is negative.

In this example, we use the illness-death model that can be used to study the effect of binary time-dependent covariates, shown in Figure 2.

The heart patients are in the illness status when the disease is diagnosed. At any time thereafter, they may become transplanted or deceased.

We propose a model involving 2 stages (Illness and Transplant) as shown in Figure 2. In each stage, the patients may move from left to right in the figure as they age. After a heart transplant, patients move right from a state in the illness stage to a state in the transplant stage. Of course, patients may die while in any state, as indicated by

the arrows to death state in Figure 2.

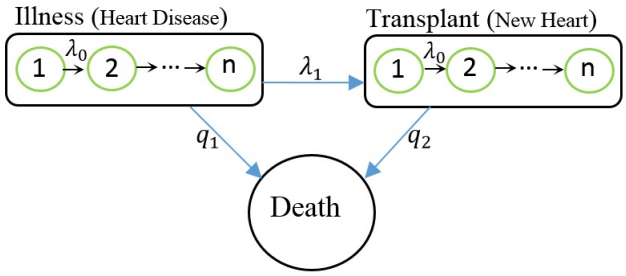

Figure 2: Illness-death model with disease and transplant states.

In order to describe how probability values are calculated, we define some notations for our model. For $t \geq 0$, let J_t represent the status of a patient at time t. We use the finite-state continuous-time Markov process with $2n$ alive states: n transient states for the disease stage, n transient states for the transplant stage, and one absorbing state which is the death state. Suppose that the space state is partitioned into the set E_1 of n illness states and the set E_2 of n transplant states and D, for death.
Hence the sub-intensity matrix (4) will be reduced to

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_{12} \\ \mathbf{0} & \mathbf{T}_2 \end{bmatrix}. \tag{11}$$

Where the $\mathbf{T}_1$ and $\mathbf{T}_2$ are the sub-intensity (with non-zero main and upper diagonal elements) matrices for the Markov chain describing a sojourn in E_1, and E_2, respectively. The matrix $\mathbf{T}_{12}$ contains the transition rates from the illness stage to the transplant stage. Let $\boldsymbol{\alpha}$ be an 2n-dimensional column vector providing the initial state probabilities.

The elements of the proposed sub-intensity matrix $\mathbf{T}$ are:

$$\mathbf{T} = \begin{bmatrix} -(q_1^1 + \lambda_0 + \lambda_1) & \lambda_0 & \cdots & 0 & \lambda_1 & 0 & 0 & \cdots \\ 0 & -(q_2^1 + \lambda_0 + \lambda_1) & \lambda_0 & \cdots & 0 & \lambda_1 & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -(q_n^1 + \lambda_1) & 0 & 0 & \cdots & \lambda_1 \\ 0 & 0 & \cdots & 0 & -(q_1^2 + \lambda_0) & \lambda_0 & 0 & \cdots \\ 0 & 0 & \cdots & 0 & 0 & -(q_2^2 + \lambda_0) & \lambda_0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & -q_n^2 \end{bmatrix}, \tag{12}$$

where λ_0 is the rate of transition from one state to the next, λ_1 is the rate of transplant from disease. The rate of mortality for the stages of decease and transplant are shown

by $\mathbf{q}_i^1$ and $\mathbf{q}_i^2$, respectively.

$$\begin{aligned} \mathbf{q}_i^1 &= a + b + q \times i^{(p+\gamma_1 age+\gamma_2 year+\gamma_3 surgery)}, \\ \mathbf{q}_i^2 &= a + q \times i^{(p+\gamma_1 age+\gamma_2 year+\gamma_3 surgery)}, \end{aligned} \tag{13}$$

for $i = 1, 2, \ldots, n$. Where the constant a is interpreted as a background rate that a general reflection of the living environment, q is a scale parameter and p is a measure of the relative impact of the occurrence of aging. By descriptive statistics from heart data, rate of death for heart patients, before transplant is higher than after. So we add the parameter b to the elements of $\mathbf{q}_i^1$. The three regression coefficients, γ_1, γ_2, and γ_3 are included in the model to represent Age, Year and Surgery effects.

Also it was assumed that patients at acceptance will be in the first phase of the model. That is, $\boldsymbol{\alpha} = (1, 0, ..., 0)'$.

We aim to estimate the parameters by the maximum likelihood method. In the heart transplant data, there are four situations for a patient: He stays in disease stage (stage 1) until the end of the study period, dies before the transplant, dies after the transplant or he is alive after transplant (stage 2) until end of the follow-up. Since the Markov process has only a single absorbing state, the time of death (the time of absorbing) follows a PH distribution. In order to estimate the parameters of the model, by using the phase-type density function (3) with representation $(\boldsymbol{\alpha}, \mathbf{T})$, we can construct expressions for the contribution of the likelihood function.

1. The probability of staying in disease state for a period of time t:

$$f_1 = Pr[J_u \in E_1 \ \ \forall\, u \in [0, t] \mid J_0 \in E_1] = \hat{\boldsymbol{\alpha}}_1' e^{\mathbf{T}_1 t}\mathbf{1}. \tag{14}$$

2. The probability of staying in disease state for s time units and then staying in transplant state until time t, will be:

$$f_{12} = \underset{\forall\, u\in[0,s] \ \ \forall\, v\in[0,t-s]}{Pr[J_u \in E_1, J_{v+s} \in E_2 \mid J_0 \in E_1]} = \hat{\boldsymbol{\alpha}}_1' \, e^{\mathbf{T}_1 s}\mathbf{T}_{12}e^{\mathbf{T}_2(t-s)}\mathbf{1}. \tag{15}$$

3. The probability of death in disease state after t time units is:

$$f_{10} = \underset{\forall\, u\in[0,t)}{Pr[J_u \in E_1, J_t \in D \mid J_0 \in E_1]} = \hat{\boldsymbol{\alpha}}_1' e^{\mathbf{T}_1 t}\mathbf{t}_{10}. \tag{16}$$

where $\mathbf{t}_{10} = -\mathbf{T}_1\mathbf{1}$.

4. The probability of death at time t from the transplant after staying in disease state for s time units is:

$$\begin{aligned} f_{20} &= \underset{\forall\, u\in[0,s] \ \ \forall\, v\in[0,t-s]}{Pr[J_u \in E_1 \, J_{v+s} \in E_2, J_t \in D \mid J_0 \in E_1]} \\ &= \boldsymbol{\alpha}_1' \, e^{\mathbf{T}_1 s}\mathbf{T}_{12}e^{\mathbf{T}_2(t-s)}\mathbf{t}_{20}, \end{aligned} \tag{17}$$

where $\mathbf{t}_{20} = -\mathbf{T}_2\mathbf{1}$. We estimate the parameters in (13) by maximizing the likelihood function,

$$L(\theta) = f_1^{n_1} f_{12}^{n_2} f_{10}^{n_3} f_{20}^{n_4},$$

where n_ks, $k = 1, 2, 3, 4$, represent the number of patients in each scenario explained in (14)-(17) and $\theta = (a,\ b,\ q,\ p,\ \lambda_0,\ \lambda_1,\ \gamma_1,\ \gamma_2,\ \gamma_3)$.

The FMINCON function in MATLAB Software has been used for finding the maximum likelihood estimates. We estimated the parameters for different numbers of n. Based on the results shown in Table 5, $n = 3$ was chosen. Note that our criterion is maximizing $l(\theta) = log(L(\theta))$.

Table 5: Parameters values for different n values.

n	a	q	p	λ_0	λ_1	b	γ_1	γ_2	γ_3	$l(\theta)$
1	1.7e-03	5.0e-06	7.5	1.50	0.0116	0.0033	1.0	-0.25	-0.54	-896.48
2	1.7e-03	6.7e-08	3.0	0.50	0.0116	0.0034	0.37	-0.17	-0.59	-896.45
3	4.9e-04	6.4e-08	9.3	0.50	0.0115	0.0034	0.098	-0.02	-0.92	-885.17
4	7.7e-04	2.7e-08	7.7	0.49	0.0113	0.0034	0.104	-0.003	-0.98	-885.88
5	7.7e-04	1.2e-08	7.1	0.49	0.0115	0.0036	0.094	-1.2e-07	-0.98	-886.25
6	8.4e-04	8.7e-09	6.5	0.49	0.0116	0.0037	0.089	-0.015	-0.87	-886.78
8	1.0e-03	5.9e-10	6.7	0.30	0.0116	0.0037	0.094	-2.1e-08	-0.86	-888.37
10	8.9e-04	5.1e-09	5.2	0.48	0.0119	0.0038	0.077	-1.3e-08	-0.75	-887.63
20	1.7e-03	7.5e-09	2.2	0.008	0.0116	0.0034	0.49	-0.47	-0.60	-895.97

The parameter values, the standard deviations and the confidence intervals at level 95% are shown in Table 6.

In the fitted model, the influence of Age at acceptance on hazard is positive, while effects of Year and Surgery are both negative as they were expected.

The standard deviation of the parameter is calculated by the bootstrap technique. The main benefit of the bootstrap is that it allows statisticians to set the confidence intervals of the parameters without having to make unreasonable assumptions. It creates multiple resamples (with replacement) from a single set of observations, and computes the effect size of interest on each of these resamples. The bootstrap resamples of the effect size can then be used to determine the 95% confidence interval. See Efron and Tibshirani (1993) for more about bootstrap techniques.

Table 6: Parameters values with standard deviations and confidence intervals (95%).

Param.	a	q	p	λ_0	λ_1	b	γ_1	γ_2	γ_3	$sup_\theta l(\theta)$
Estimate	4.9e-04	6.4e-08	9.3	0.50	0.0115	0.0034	0.098	-0.02	-0.92	-885.17
Std.	5.3e-04	2.9e-06	2.36	0.34	0.0036	0.0018	0.49	0.37	0.22	
(Lower,)	1.9e-04	9.2e-14	1.5e-08	4.8e-04	0.007	4.9e-04	0.055	-0.50	-0.98	
(,Upper)	0.0025	1.0e-05	9.7	1.25	0.021	0.0077	1.77	-1.8e-08	-0.0014	

It was our interest to know about the significance of the covariates and the parameter

b in the model. To this end, we have conducted hypothesis testing for some null tests. In Table 7 the results are presented.

Table 7: The value of estimates

Parameter	a	q	p	λ_0	λ_1	b	γ_1	γ_2	γ_3	$sup_\theta l(\theta)$
without γ_1,γ_2,γ_3	0.0017	4.3e-09	3.01	0.50	0.0116	0.0033	-	-	-	-896.48
without γ_1	0.0013	9.9e-06	5.1	0.50	0.0114	0.0029	-	-0.50	-0.96	-895.54
without γ_2	6.1e-04	1.2e-08	10.4	0.21	0.0114	0.0036	0.14	-	-0.98	-885.02
without γ_3	5.0e-04	3.5e-07	7.4	0.50	0.0116	0.0036	0.13	-3.7e-04	-	-888.01
without b	0.0022	2.1e-09	0.39	0.0035	0.0112	-	1.71	-0.077	-0.52	-904.7

The likelihood ratio tests (LRT) have been used for the testing. For example, for the null hypothesis $H_0 : \forall\, i, \gamma_i = 0$ against the alternative $H_1 : \exists\, i, \gamma_i \neq 0$ can be expressed:

$$LRT = -2log\left(\frac{sup_{\Theta_0} L(\theta)}{sup_{\Theta} L(\theta)}\right) = -2(sup_{\Theta_0} l(\theta) - sup_{\Theta} l(\theta)) = 22.62, \tag{18}$$

which is greater than $\chi^2_{0.95,3} = 7.8$ and therefore the null hypothesis is disproved. As seen in Table 7, we can obtain the values of LRT for the model without every variable separately. The parameter space and the parameter under the null hypothesis are Θ and Θ_0 respectively.
For testing the null hypothesis $H_0 : b = 0$ against the alternative $H_1 : b \neq 0$, LRT is 39.06, by comparing with $\chi^2_{0.95,1} = 3.8$, we reject H_0. Hence, it can be said that mortality before heart transplant is greater than after transplant.

Now using the value of parameters, the probability of the number of transition until time t can be calculated. These probabilities are shown in Table 8 for different Ages (30 and 50), Years (years 3 and 5) and Surgery (0 and 1). As we can see in this table, the probability without transition decreases as t increases because the patient may die or transplant after more time of staying in disease stage. The event of no transition decreases with age, and increases with year and surgery. For example, a 30-year-old patient who was diagnosed a heart disease in the third year of the study without heart surgery, the probability of staying in the same state after one month is equal to 0.6254, and in While this probability is 0.5955 for a 50-year-old one. So increasing age has a decreasing effect on this possibility. Mortality decreases with year and surgery, so the probability for no transition also increases as the year and surgery increase.

Also mortality increases with age, so the probability of staying in the disease stage for older patients is reduced, and he/she is more likely to die. Therefore, it is expected that the probability of one transition increases with age. The probability of one transition increases until 6 months but decreases thereafter because the probability of transplant decreases in the later months and the probability of death increases. Also, in a more detail, this probability, with fixed year and surgery by increasing age, increases until one month and decreases after that. For example, a 30-year-old patient who diagnose a heart disease in the third year of the study without heart surgery, the probability of one transitions after one month is equal to 0.3714, and in While this probability is

0.3935 for a 50-year-old one. They are respectively 0.7338 and 0.7215 for 30 and 50 year-old after 3 months.

The probability of two transitions constantly increases. It is obvious that this probability increases with age because of its positive influence on death, decreases with year and surgery because of their negative influence on death. For example, a 30-year-old patient who diagnose a heart disease in the third year of the study without heart surgery, the probability of two transitions after three months is equal to 0.0221, and in While this probability is 0.0709 for a 50-year-old one. So this probability by increasing age, increase. Now with the same condition, this probability for a patient who was diagnosed in the fifth year of the study is 0.0175. So we can say that by increasing the year of acceptance, this probability decreases and this result also applies to surgery.

Table 8: Probability of N(t) for different Age, Year and Surgery.

Age	Year	Surgery	P[N(t)]	1 month	3 months	6 months	1 year	3 years
30	3	0	P[N(t)=0]	0.6254	0.2441	0.0595	0.0033	3.5e-08
			P[N(t)=1]	0.3714	0.7338	0.8786	0.8505	0.6082
			P[N(t)=2]	0.0032	0.0221	0.0619	0.1462	0.3918
		1	P[N(t)=0]	0.6279	0.2474	0.0612	0.0035	4.1e-08
			P[N(t)=1]	0.3695	0.7350	0.8892	0.8776	0.6650
			P[N(t)=2]	0.0026	0.0176	0.0496	0.1190	0.3350
	5	0	P[N(t)=0]	0.6255	0.2443	0.0596	0.0033	3.5e-08
			P[N(t)=1]	0.3713	0.7339	0.8793	0.8523	0.6117
			P[N(t)=2]	0.0032	0.0218	0.0611	0.1444	0.3883
		1	P[N(t)=0]	0.6280	0.2475	0.0612	0.0035	4.2e-08
			P[N(t)=1]	0.3695	0.7350	0.8894	0.8783	0.6666
			P[N(t)=2]	0.0026	0.0175	0.0493	0.1183	0.3334
50	3	0	P[N(t)=0]	0.5955	0.2077	0.0428	0.0017	4.5e-09
			P[N(t)=1]	0.3935	0.7215	0.7764	0.6360	0.3831
			P[N(t)=2]	0.0110	0.0709	0.1808	0.3624	0.6169
		1	P[N(t)=0]	0.6168	0.2332	0.0542	0.0027	2.0e-08
			P[N(t)=1]	0.3777	0.7299	0.8453	0.7712	0.4802
			P[N(t)=2]	0.0055	0.0369	0.1005	0.2261	0.5198
	5	0	P[N(t)=0]	0.5969	0.2093	0.0434	0.0017	5.0e-09
			P[N(t)=1]	0.3925	0.7220	0.7804	0.6426	0.3844
			P[N(t)=2]	0.0106	0.0688	0.1762	0.3557	0.6156
		1	P[N(t)=0]	0.6173	0.2339	0.0545	0.0027	2.0e-08
			P[N(t)=1]	0.3774	0.7301	0.8472	0.7756	0.4859
			P[N(t)=2]	0.0053	0.0360	0.0982	0.2217	0.5141

Given $J_u \in E_i$, the time of staying continuously in stage i of disease has a PH distribution with representation $(\hat{\boldsymbol{\alpha}}_i^u, \mathbf{T}_i)$, where

$$\hat{\boldsymbol{\alpha}}_i^u = \frac{\hat{\boldsymbol{\alpha}}' \, e^{\mathbf{T}u} \mathbf{I}_{E E_i}}{\hat{\boldsymbol{\alpha}}' \, e^{\mathbf{T}u} \mathbf{1}_i},$$

$\mathbf{I}_{E E_c}$, $c = 1, \ldots, k$, is $nk \times n$ matrix with $(n(c-1)+l, l)$ entry equal to 1, $l = 1, \ldots, n$ and all other entries are 0, and $\mathbf{1}_c$ is a column vector with nk entries such that elements $n(c-1)+l$, $l = 1, \ldots, n$ equal to 1 and all other entries 0.

Hence, given that the process J is i^{th} stage at time u, the expected time of continuously staying in this stage, in time interval $[0, t]$ is equal to:

$$
\begin{aligned}
E\left[\int_0^\infty \mathbf{1}_{\{J_{u+s}\in E_i,\,\forall\, s\in[0,t]\}} ds \mid J_u \in E_i\right] &= \int_0^\infty Pr[J_{u+s} \in E_i,\, \forall\, s \in [0,t] \mid J_u \in E_i]\, ds \\
&= \int_0^t \hat{\boldsymbol{\alpha}}_i^u\, e^{\mathbf{T}_i s} \mathbf{1}_i\, ds \\
&= \hat{\boldsymbol{\alpha}}_i^u\, \mathbf{T}_i^{-1}(e^{\mathbf{T}_i t} - \mathbf{I})\mathbf{1}_i,
\end{aligned} \tag{19}
$$

where the time of staying in stage i has a PH distribution with representation of $(\hat{\boldsymbol{\alpha}}_i^u, \mathbf{T}_i)$. The second equality holds due to PH property; in fact, the integrand is the survival function of a PH distribution. The matrix $\mathbf{I}$ is an n-dimensional identity matrix and $\mathbf{1}_i$ is an n-dimensional column vector. Given starting from stage illness (stage 1), expected times without transition for $u = 0$, are shown in Table 9 for different ages. It is expected that the 30-year-old patient will spend 24 days in the disease stage

Table 9: Expected sojourn time (days) of without transition

	duration				
Age	1 month	3 months	6 months	1 year	3 years
30	24	48.3	60.1	63.6	63.9
40	23.9	47.6	58.7	61.9	62.1
50	23.5	45.6	55.0	57.3	57.4
60	22.5	40.5	46.3	47.3	47.3

Finally, the probability of transition from stage i to stage j in a period of time t can be calculated easily as follows

$$
Pr[J_{u+t} \in E_j \mid J_u \in E_i] = \frac{\hat{\boldsymbol{\alpha}}' e^{\mathbf{T}u} \mathbf{I}_{E\,E_i} \mathbf{I}'_{E\,E_i} e^{\mathbf{T}t} \mathbf{1}_j}{\hat{\boldsymbol{\alpha}}' e^{\mathbf{T}u} \mathbf{1}_i}. \tag{20}
$$

For Age= 30, Year= 3 and Surgery= 0, and $u = 0$ (20) can be seen in Table 10. The formula for the probability of transition from disease stage to transplant is greater than death. Mortality from disease and transplant stages increase over time but mortality before transplant is greater than after.

Table 10: Probability of transition from stage i to stage j for Age=30, Year=3, Surgery=0.

from	to	1 months	3 months	6 months	1 year	3 years
disease	transplant	0.2726	0.5338	0.6296	0.5866	0.3434
disease	death	0.0988	0.2000	0.2490	0.2640	0.2648
transplant	death	0.0032	0.0221	0.0619	0.1462	0.3918

Example 2: Cancer Disease
After someone is diagnosed with cancer, doctors will try to figure out if it has spread, and if so, how far. This process is called staging. The stage of cancer describes how much cancer is in the body. It helps to determine how serious the cancer is and how best to treat it. Cancer staging may sometimes include the grading of the cancer. This describes how similar a cancer cell is to a normal cell. Doctors also use a cancer's stage when talking about survival statistics. The earliest stage of cancer is called stage 0 (carcinoma in situ), and then ranges from stages I (1) through IV (4). As a rule, the lower the number, the less the cancer has spread. A higher number, such as stage IV, means cancer has spread more. Although each person's cancer experience is unique, cancers with similar stages tend to have a similar outlook and are often treated in much the same way. When cancer returns after a period of remission, it's considered a recurrence. A cancer recurrence happens because, in spite of the best efforts to rid of cancer, some cells from cancer remain. These cells could be in the same place where cancer first originated, or they could be in another part of body. These cancer cells may have been dormant for a period of time, but eventually they continued to multiply, resulting in the reappearance of the cancer.

Also by the end of the observation period, the patient under study may not have reached an absorbing state. In survival analysis this would correspond to an individual still being alive by the end of the study and this kind of incomplete observation is known as right-censoring.

To illustrate the computation of the above probabilities, we use all transition functions described in section 3. In the following, we show how the PH model can be used in modeling of recurrent events in cancers.

We assume 5 disease stages and one recovery stage, each including 5 states, *i.e.* $k = 6$ and $n = 5$. Stage 1 corresponds to the recovery (R) and stages $2, 3, \ldots, 6$ represent cancer stages $0, 1, \ldots, 4$, respectively. The entries of the sub-generator matrix $\mathbf{T} = (t_{ij})$ $i, j = 1, ..., 30$ will be described in the following.
The rates of transition from one state to the next are given by

$$t_{i,i+1} = \lambda, \quad i = 1, 2, ..., 29, i \neq 5, 10, 15, 20, 25,$$

where λ is assumed to be 0.2.

The rates of recovery from disease stage, defined based on the degree of malignancy of the disease, are

$$t_{i+5l,i-5} = \gamma(0.1)^l, \quad i = 6, 7, ..., 10, \ l = 0, 1, ..., 4,$$

where γ is the coefficient of recovery; assume $\gamma = 0.1$ and the rates of transition from one stage of disease to the stage before, defined based on the degree of malignancy of the disease, are given by

$$t_{i+5l,i+5(l-1)} = \beta(0.1)^l, \quad i = 11, 12, ..., 15, \ l = 0, ..., 3,$$

where β is assumed to be 0.2.

The rates of transition from one stage of disease to the next are

$$t_{i+5l,i+5(l+1)} = (l+1)(a+qi^p), \quad i = 1,2,...,5,\ l = 1,...,4,$$

with $a = 10^{-3}$, $q = 10^{-6}$ and $p = 4.5$.

And the rates of transition from recovery stage to each of the disease stages are given by

$$t_{i,i+5(l+1)} = (0.1)^l(a+qi^p), \quad i = 1,2,...,5,\ l = 0,1,...,4.$$

And the mortality rates are given by

$$\mathbf{t}_0' = (t_{1,0}, t_{2,0}, ..., t_{30,0}),$$

where

$$t_{i,0} = (0.1)^5 + a + qi^p, \quad i = 1,...,5,$$

and for $l = 0,1,...,4$

$$t_{i+5(l+1),0} = (0.1)^{4-l} + a + qi^p, \quad i = 1,...,5.$$

So other entries of **T** are zero except the diagonal entries which are:

$$t_{i,i} = -\sum_{j=0,j\neq i}^{30} t_{i,j}, \quad i = 1,2,...,30.$$

Finally, the initial probability is $\boldsymbol{\alpha}' = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2', ..., \boldsymbol{\alpha}_6')$ where $\boldsymbol{\alpha}_i$'s are 5-dimensional column vector. After the diagnosis of cancer, first stage, i^{th}, is also known so $\boldsymbol{\alpha}_i'$ is $(1,0,0,0,0)$ and other elements of $\boldsymbol{\alpha}$ will be zero.

Table 11: Probability of N(t) in Example 2.

		duration			
	input stage	6 months	12 months	24 months	36 months
P(N(t)=0)					
	0	0.5382	0.2885	0.0814	0.0226
	1	0.2750	0.0752	0.0055	0.0004
	2	0.8046	0.6429	0.3981	0.2403
	3	0.5219	0.2701	0.0698	0.0175
	4	0.0025	6.04e-6	3.62e-11	2.16e-16
P(N(t)=1)					
	0	0.4541	0.6880	0.8504	0.8576
	1	0.5199	0.4422	0.1960	0.0952
	2	0.1406	0.2061	0.2694	0.2959
	3	0.4573	0.6915	0.8703	0.9132
	4	0.9975	0.9999	0.9998	0.9998
P(N(t)=2)					
	0	0.0088	0.0204	0.0440	0.0669
	1	0.2875	0.4614	0.3268	0.1523
	2	0.0639	0.1173	0.1483	0.1552
	3	0.0208	0.0377	0.0559	0.0622
	4	7.94e-5	1.31e-4	1.71e-4	1.81e-4

Table 11 shows the probability of $N(t)$, (5)–(7), for different input stages, for different durations. These probabilities behave the way we expect. With entering to every stage, the probability of no transition decreases as the duration increases. This means that probability of staying in every stage is higher at the earlier times and decreases over time. For example, in stage 4, probability is near zero over time, because staying in this state is very low and the patient has a higher chance of dying. For one transition until t this probability is near to one, because of the transition to death state. As it is seen in Table 11, the probability of staying in stage 2 is the maximum, compared to other stages, due to the fact that chance of recovery is not as high as in stage 1 and the rate of death is not as high as in stage 3.

The expected time of staying in stage at diagnosis, (19), is shown in Table 12. When a patient is diagnosed with stage 0 cancer, it is expected that he stays in that stage for 4.5 months during the six months. If the first stage is 2, this value will be 5.4 months. As we said before, at this stage, the expected sojourn time is greater than the others because of the low chance of recovery compared to the previous stage and low chance of death compared to the next stage. The patient, who is in stage 4, is expected to eventually survive for 1 month.

Table 12: Expected sojourn time in first stage until t.

	duration			
input stage	6 months	12 months	24 months	36 months
0	4.5	6.9	8.8	9.4
1	3.4	4.3	4.6	4.6
2	5.4	9.7	15.9	19.6
3	4.4	6.7	8.5	9.0
4	1	1	1	1

Also we can obtain the probability of one transition from stage i to stage j, $P[N_{ij}(t) = 1]$, by using Equation (20).

Table 13: Probability of one transition between stages.

stage	waiting time	R	0	1	2	3	4	D
0	6 months	0.4440	0	0.0050	0	0	0	0.0051
	12 months	0.6750	0	0.0046	0	0	0	0.0084
1	6 months	0.0334	0.4704	0	0.0092	0	0	0.0069
	12 months	0.0420	0.3809	0	0.0104	0	0	0.0089
2	6 months	0.0054	0	0.0592	0	0.0165	0	0.0596
	12 months	0.0096	0	0.0635	0	0.0247	0	0.1084
3	6 months	4.38e-04	0	0	0.0078	0	0.0032	0.4459
	12 months	6.58e-04	0	0	0.0103	0	0.0021	0.6784
4	6 months	9.89e-6	0	0	0	1.16e-4	0	0.9973
	12 months	9.72e-6	0	0	0	6.049e-5	0	0.9998

As seen in Table 13, probability of one transition from the lower stages of cancer to recovery stage is higher than the higher stages, this means that the lower the number, the less the cancer has spread and the higher number of cancer has spread more. In higher stages probability of transition to recovery is near to zero. Also probability of

transition from the lower stages of cancer to death is lower than the higher stages and probability of transition from higher stages to death is near to one. Other possible probabilities from one stage to another are shown in this table.

4 Conclusion

In this article, we work on data with recurrent events that are widely used in the medical studies for many diseases. The most common of which are surgeries and cancers. It is important for patients to be aware of information about their illness, such as the probability of relapse, the time stay in each stage of recovery or disease, the probability of recovery, and so on. Using Markov's properties and PH distributions, we present a formula for calculating the probability of the number of occurrences of these events, which are interdependent and defined as recursive relations, and we need to use differential equations to calculate them. In this article, we used two examples to introduce this formula and get interpretable results. Example 1 of the real data was for 103 Stanford heart patients who have two stages: heart disease and heart transplant. The mortality rate was defined for these stages, which have several parameters. We estimate these parameters by the maximum likelihood method to have the matrix $\mathbf{T}$. A vector of covariates including Age, Year, previous Surgery are estimated where the influence of Age on hazard is positive, while effects of Year and Surgery are both negative and mortality before heart transplant is greater than after transplant. The standard deviations and the confidence intervals are obtained by the bootstrap technique. Finally, the probability of the number of transition until time t, the expected time of continuously staying in each stage and the probability of transition from stage i to stage j in a period of time t are obtained. The results are expected as shown in the tables. the probability without transition decreases as t increases because the patient may die or transplant after more time of staying in disease stage. The probability of one transition increases until 6 months but decreases thereafter because the probability of transplant decreases in the later months and the probability of death increases. The probability of two transitions constantly increases. Another example is a simulated cancer that has one recovery stage and five cancer stages. In the simulated example, which had a more complex model, similar calculations have been done.

Calculating these probabilities for higher recurrent numbers is time consuming due to more transition between stages. Because of using of the recursive equations in the formulas, we need to use the differential equations of the previous equations to calculate the probability. In other words, we have to start from the beginning to get the probability.

We have used the available functions in Matlab and have not developed any algorithms for solving the differential equations. In future studies, we aim to optimize the algorithms for calculating probability distribution of the transitions.

References

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer, New York.

Andersen, P. K. and Gill, R. D. (1982), Cox's regression model for counting processes: a large sample study. *Ann Stat*. **10**, 1100–20.

Araujo, A., Meira-Machado, L. and Roca-Pardinas, J. (2014). TPmsm: Estimation of the Transition Probabilities in 3-State Models. *Journal of Statistical Software*, **62**(4), 1–29.

Asghari, R. and Hassan Zadeh, A. (2020). Mortality modeling of skin cancer patients with actuarial applications. *Journal of North American Actuarial*, 0(0), 1–17, DOI:10.1080/10920277.2019.1670070.

Asmussen, S., Nerman, O. and Olsson, M. (1996). Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, **23**(4), 419–441.

Balboa, V. and Una-Alvarez, J. (2018). Estimation of Transition Probabilities for the Illness-Death Model: Package TP.idm. *Journal of Statistical Software*, **83**(10), 1–19.

Cook, R. J. and Lawless, J. F. (2010). *The Statistical analysis of recurrent events*, 2nd edn., New York, NY, Springer.

Crowley, J. and Hu, M. (1977). Covariance Analysis of Heart Transplant Survival Data, *Journal of the American Statistical Association*, **72**, 27–36.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*, New York: Chapman & Hall.

Faddy, M. J. and McClean, S. I. (2000). Analysing data on lengths of stay of hospital patients using phase-type distributions. *Applied Stochastic Models in Business and Industry*, **15**(4), 311-317.

Hassan Zadeh, A., Jones, B.L. and Stanford, D. A. (2013). The use of phase-type models for disability insurance calculations. *Scandinavian Actuarial Journal*, **2014**, 714–728.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*, Springer.

Kalbfleisch, J. D. and Prentice RL. (1980). *The Statistical Analysis of Failure Time Data*, John Wiley & Sons.

Kelly, P.J. and Lim, L. LY (2000). Survival analysis for recurrent event data: an application to childhood infectious diseases. *Stat Med*, **19**, 13–33.

Kleinbaum, D.G. and Klein, M. (2012). *Survival Analysis: A Self-Learning Text*, Third Edition, Springer.

Latouche, G., Ramaswami, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling*, Series on statistics and applied probability.

Lin, X. Sh. and Liu, X. (2007). Markov aging Process and Phase-Type Law of Mortality. *North American Actuarial Journal*, **11**(4), 92–109.

Lin, D. Y., Wei, L. J., Yang I., Ying Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J R Stat Soc B*, **62**, 711–30.

Marshall, A. and Zenga, M. (2009). Simulating coxian phase-type distributions for patient survival. *International Transactions in Operational Research*, **16**(2), 213–226.

Meira-Machado, L., Una-Alvarez, J., Cadarso-Suarez, C., Andersen, P. K. (2009). Multi-state models for the analysis of time to event data. *Statistical Methods in Medical Research*, **18**, 195–222.

Moreira, A. C. (2013). *Methods for Analysis of Mult-state Survival Data*, Doctoral dissertations, University of Minho.

Neuts, M. F. (1994). *Matrix-geometric solutions in stochastic models: An algorithmic approach*, Mineola NY: Dover Publications.

Aalen O. O. (1995). Phase type distributions in survival analysis. *Scandinavian Journal of Statistics*, **22**(4), 447-463.

Reinecke P., Bodrog L., Danilkina A. (2012). *Phase-type distributions; In Resilience Assessment and Evaluation of Computing Systems*, Springer Berlin Heidelberg.