Abstract

Multi-state models are powerful tools to understand and describe complex disease. Patients can move among a certain number of states defined by specific conditions of disease level often including death. Typically in these studies, the issues of interest include overall survival, effects of prognostic factors on disease progress and estimation of transition probabilities. Often multi-state models are employed under the Markov assumption, or it is assumed that the multi-state model can be described by a Markov renewal process. These assumptions are mainly made for mathematical convenience, since it is easier to estimate transition intensities and covariate effects. Moreover the transition probabilities can be computed. Obviously these assumptions could be rather unrealistic or too restrictive. The Markov assumption might not hold because there is association between transition times. In case a positive association is present, later transition will show higher rate if earlier transitions had taken place earlier. This implies a violation of the Markov assumption because the future depends not only on the present status but also on the past.

In this thesis the Markov Renewal assumption is relaxed and two methods are proposed to deal with a violation of this assumption. The first method focus on the illness-death model, which is a 3-states model where only an intermediate event can occur before the main event of interest takes place. By relaxing the Markov renewal assumption, the proposed method models the correlation between transition times in the framework of Cox model. To obtain predictions for patients with a given history, formulas for prediction of transition probabilities are developed. For application purpose, some general functions for prediction are also developed in R and their use is illustrated through a set of data coming from a breast cancer trial.

Relying on the frailty theory, the second approach proposed in this thesis models a forward-going sequential process in a framework of hidden Markov model. By extending the two-point mixture frailty model, frailties are modeled as hidden states which can have an impact on the transition rates and eventually be observed by the sojourn times and occurrence of events. Based on the likelihood construction an Expectation-Maximization algorithm was proposed.
Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Marta Fiocco for her encouragement, positive attitude and advice during the progression of this thesis. This thesis would not have been possible without her patient guidance, even at late night or during summer vacation. I want to give special thank to Professor Jacqueline Meulman for her kindly support for the extension of my academic year registration. I want to thank Professor Hein Putter for his helpful advice. I want to thank our survival group for helpful comments and suggestion. The European Organization for Research and Treatment of Cancer (EORTC) is gratefully acknowledged for providing the data. I am grateful to my professors from the Statistical Science for the Life and Behavioural Sciences master track for the two years of inspiration that they offered me. I want to thank all my fellow students with whom I shared laughs and cries over assignments and exams. I wish to thank my family and friend for their support and pretending to be interested on my talking during the progress of this thesis.
Contents

1 Introduction
   1.1 Aim of the Thesis ........................................... 6
   1.2 Structure of the Thesis .................................... 6

2 Introduction to Survival Analysis and Multi-state Model ..... 7
   2.1 General Concepts .......................................... 7
   2.2 General Concept about Multi-state Models ................. 8
   2.3 Time Scale in Multi-state Model ........................... 8
   2.4 Cox Proportional Hazard Model ............................ 9

3 Markov Models .................................................. 11
   3.1 Markov Assumption ........................................ 11
   3.2 Markov renewal Assumption ............................... 11

4 Thesis Contribution: Extended Markov renewal Model ......... 12
   4.1 Estimation of Parameters .................................. 12
   4.2 Prediction Formulas ...................................... 13
   4.3 Standard Error of Prediction ............................. 14

5 Application ...................................................... 17
   5.1 Data Description .......................................... 17
   5.2 Formats of Presenting Data ................................ 18
   5.3 Data Analysis ............................................. 19

6 Software for Prediction in Extended Markov Renewal Model ... 22
   6.1 Implementation for Prediction Formulas .................. 22
   6.2 Implementation for Simulation Method .................... 25

7 Results and Conclusions in the Extended Markov Renewal Model 28

8 Frailty Model .................................................. 31
   8.1 Frailties in simple survival models ........................ 31
   8.2 Two-point Mixture Frailty Distribution .................... 32

9 Hidden Markov Model .......................................... 33

10 Thesis Contribution: Hidden Markov Two Point Frailty Model 34
   10.1 Definitions and Notations .................................. 34
   10.2 Hidden Markov: extension for the two-points frailties .... 35

11 Work in Progress: EM Algorithm ................................ 37
1 Introduction

Multi-state models are widely used for describing the longitudinal progression of subjects between a finite number of states, who are exposed to several time-dependent stochastic events. Occurrence of an event, resulting in change of state, is called transition. In medical application, multi-state models are useful tools for modeling the disease process of patients when intermediate events of interest can occur during the follow-up time [1].

This class of models can allow much flexibility when interests are in multiple survival outcomes. In a breast cancer trial, for instance, intermediate events like recurrence of tumor in the vicinity of the primary tumor (local recurrence), or at distant locations (distant metastasis) occur after surgery of the primary tumor. Figure 1 shows an example multi-state model for such trial. A patient can experience relapse-free survival, local recurrence, distant metastasis or both after the surgery, and may eventually end up in the absorbing state death. Clinicians might be interested in evaluating influence of several treatments on both overall survival and occurrence of the intermediate events. In these situations, multi-state model can be used to model patient’s history, and evaluate the influence of prognostic factors on possible transitions between the states. It is also valuable to predict probabilities of visiting a certain state within a given time for a patient with specific clinical prognosis by employing such models [2].

![Figure 1: An example multi-state model for breast cancer trial](image)

---

[1] Reference 1
[2] Reference 2
1.1 Aim of the Thesis

Often when employing a multi-state model, the Markov assumption is adopted to simplify the inference process. The Markov property states that the future depends on the history only through the present. For a multi-state model this means that, given the present state and the event history of a patient, the next state to be visited and the time at which this will occur will only depend on the present state. However, this assumption might fail to hold in some applications, leading to inconsistent estimates.

The goal of this thesis is to address the violation of Markov assumption in multi-state model by proposing two new methods.

The first method is an extension of Markov renewal model. Here an extra covariate to account for transition time to intermediate event is introduced. Formulas are developed to obtain the prediction for transition probabilities in an illness-death extended Markov renewal model. General functions are written in R to make the method developed in this thesis applicable to data set that can be described by an illness-death model.

The second approach combines hidden Markov model and the two-point mixture frailties. This model can deal with the association between transition times as well as the violation of Cox proportional hazard assumption. An EM algorithm is proposed to implement this method.

1.2 Structure of the Thesis

Basic concepts of survival analysis and multi-state model are described in Chapter 2. In Chapter 3, a brief introduction to Markov and Markov renewal models is given. Our first proposed method extended the Markov renewal model and it is described in Chapter 4 along with specific formulas for prediction and the simulation algorithm. In Chapter 5 the model is applied to a breast cancer data set. Details concerning the softwares written to implement the proposed method are given in Chapter 6. The results and conclusions concerning the application of the extended Markov renewal model to the breast cancer data set are outlined in Chapter 7.

The second part of this thesis concerns frailty and hidden Markov model. A short introduction to frailty model and hidden Markov model is given in Chapter 8 and 9 respectively. To deal with possible association between transitions a hidden Markov two-point frailty model is proposed in Chapter 10 and an EM algorithm is outlined in Chapter 11. All R code written for this thesis can be found in the appendices.
2 Introduction to Survival Analysis and Multi-state Model

In this section, a summary of the basic mathematic theory underlying survival analysis and multi-state will be presented.

2.1 General Concepts

Survival analysis studies the distribution time from an initiating state (like birth, start of treatment) to some terminal event (like death, relapse). Let $T$ be the random variable representing the interval length from the starting point to the occurrence of the event of interest. The survival function $S(t)$ represents the probability that a population survives at least until time $t$ and it is defined as

$$S(t) = P(T > t).$$

Let $F(t)$ be the cumulative distribution function of $T$, i.e. $F(t) = P(T \leq t)$. The survival function is the complement of the cumulative distribution $S(t) = 1 - F(t)$.

The survival can be also expressed as

$$S(t) = P(T > t) = \int_t^\infty f(v)dv.$$

Hazard rate function $h(t)$ expresses the rate at which an individual who is event-free at time $t$ will experience the event of interest in the next instant. It is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T > t)}{\Delta t}.$$

The cumulative hazard function is given by

$$H(t) = \int_0^t h(v)dv,$$

which is a measure of risk of the occurrence of event.

If the random variable $T$ is continuous the relation between survival and cumulative hazard is as follow

$$S(t) = \exp(-H(t)) = \exp(-\int_0^t h(v)dv).$$
2.2 General Concept about Multi-state Models

Denote by $S = \{1, ..., r\}$ the states in the multi-state model; let $X(t)$ and $H_{t-}$ be a stochastic process taking a value in $S$ at time $t$ and the history observations of the disease process over time interval $[0, t)$ respectively. The hazard function, or transition intensity, which expresses the instantaneous risk of a transition from state $g$ to state $h$ at time $t$ is defined as

$$\alpha_{gh}(t) = \lim_{\Delta t \to 0} \frac{P(X(t + \Delta t) = h | X(t) = g, H_{t-})}{\Delta t}.$$ 

The cumulative transition hazard is as follow

$$A_{gh}(t) = \int_0^t \alpha_{gh}(u)du.$$ 

If a direct transition between state $g$ and state $h$ cannot occur, $A_{gh}(t) \equiv 0$. The intensity matrix $A(t)$ can be constructed into a $r \times r$-matrix, with non-diagonal elements $A_{gh}(t) \forall g \neq h$ and diagonal elements $A_{gg}(t) = -\sum_{g \neq h} A_{gh}(t)$.

The transition probability $P_{gh}(s, u)$ is often the prime quantity of interest. It is defined as

$$P_{gh}(s, u) = P(X(u) = h | X(s) = g, H_{s-}),$$  \hspace{1cm} (1)$$

which denote transition probability of going from state $g$ to state $h$ in time interval $(s, t]$ given the patient’s history.

2.3 Time Scale in Multi-state Model

For the definition of time $t$ in the hazard functions $\alpha(t)$ defined in Section 2.2, two distinct approaches are often used in multi-state models, which are denoted here by the ‘clock forward’ or ‘clock reset’ approach.

‘Clock forward’: in this approach, time $t$ refers to the time since patient enters the initial state. That means the clock starts with 0 at entrance of the initial state and then keeps moving forward.

‘Clock reset’: in this approach, time $t$ refers to the time since patient enters the current state. That means the clock is set back to 0 each time the patient enters a new state.

The difference between the two approaches is illustrated by a cancer patient’s disease progress (Figure 2). The upper graph shows the calendar date of surgery and subsequent events. The patient is censored for death due to the end of follow-up. The lower graph compares the two different time scales. In the ‘clock forward’ approach, time is measured by years from the date of surgery, whereas in the ‘clock reset’ approach, time is measured by time intervals between occurrence of events.
Figure 2: Illustration of the ‘clock forward’ and ‘clock reset’ approach. (LR, DM and FUP stand for local recurrence, distant metastasis and follow-up, respectively.)

### 2.4 Cox Proportional Hazard Model

Cox proportional hazard model, often abbreviated to Cox model or proportional hazard model, is widely used to quantify the covariate effects on survival [3]. Under the proportional hazard assumption the effect of a unit increase in a covariate is multiplicable with respect to hazard rate. This model can be employed in multi-state model to evaluate effects of prognostic factors on different transitions. For a patient with covariate vector \( Z \), the transition-specific hazard rate transition \( g \rightarrow h \) is given by

\[
\alpha_{gh}(t) = \alpha_{gh,0}(t) \exp(\beta_{gh}^T Z)
\]

where \( \alpha_{gh,0}(t) \) is the baseline hazard of transition \( g \rightarrow h \), and \( \beta_{gh} \) is the vector of regression coefficients.

The model can be also written as

\[
\alpha_{gh}(t) = \alpha_{gh,0}(t) \exp(\beta_{gh}^T Z_{gh})
\]

where \( Z_{gh} \) is a vectors of covariates specific to transition \( g \rightarrow h \), defined for the individual based on her covariates \( Z \) [5]. Denote by \( Z_{gh,i} \) the transition-specific covariates of patient \( i \) for transition \( g \rightarrow h \). Estimates \( \hat{\beta} \) can be obtained together by maximizing the generalized Cox partial likelihood

\[
L(\beta) = \prod_{\text{transition } g \rightarrow h} \prod_{i=1}^{n} \frac{\exp(\beta_{gh,i}^T Z_{gh,i})}{\sum_{j \in R_{gh}(t_{gh,i})} \exp(\beta_{gh,j}^T Z_{gh,j})}
\]
where \( t_{gh,i} \) is the failure or censoring time of individual \( i \) for transition \( g \rightarrow h \), \( d_{gh,i} = 1 \) if individual \( i \) has an event for transition \( g \rightarrow h \), 0 otherwise, and \( R_g(t_{gh,i}) \) is the risk set of individuals who are in state \( g \) at time \( t \) (\( t \) being here the time since entry in state \( g \)). The Nelson-Aalen estimate of the cumulative baseline hazard of transition \( g \rightarrow h \) is given as follows:

\[
\hat{A}_{gh,0}(t) = \sum_{t_{gh,i} \leq t} \frac{d_{gh,i}}{\sum_{j \in R_g(t_{gh,i})} \exp(\beta^T Z_{gh,j})}
\]
3 Markov Models

3.1 Markov Assumption

In applications of multi-state models, Markov assumption is often adopted. The Markov property assumes that the future evolution of process only depends on the current state. Under the Markov assumption, the transition probability defined in (1) satisfies

\[ P_{gh}(s, u) = P(X(u) = g | X(s) = h) \]  \hspace{1cm} (2)

Markov property drastically simplifies the inference of likelihood, and under such assumption the estimation of transition probabilities can be expressed as a function of transition intensities in the form of product integral \[6\]:

\[ P(s, u) = \prod_{(s, u)} (I + dA(t)) . \]

Usually, “clock-forward” approach is used for Markov model, which means that the time scale is the calendar time since the origin of the process.

3.2 Markov renewal Assumption

If the transition intensities depend on the history not only through the current state but also on the sojourn time in the current state, the multi-state model becomes a Markov renewal model. It is also defined as semi-Markov model [7-11]. To define the Markov renewal property, we shall use the definitions and the formalism of Dabrowska et al. [10-11].

Let \( 0 = T_0 < T_1 < ... < T_m \) be consecutive times of entrance into the states \( S_0, S_1, ..., S_m \in \{1, ..., r\} \), then \((S, T) = (S_\ell, T_\ell)_{\ell \geq 0}\) forms a Markov renewal process if the sequence of states visited \( S = (S_\ell : \ell > 0) \) is a Markov chain and the sojourn times \( J_{m+1} = T_{m+1} - T_m \) satisfy:

\[ \mathbb{P}\{S_{m+1} = j, J_{m+1} \leq \tau | S_0, T_0, ..., S_m, T_m\} = \mathbb{P}\{S_{m+1} = j, J_{m+1} \leq \tau | S_m\}. \]

For Markov renewal models, “clock-reset” approach is commonly used due to the renewal nature of the process.
4 Thesis Contribution: Extended Markov renewal Model

In some applications both Markov and Markov renewal assumption could fail to hold due to the presence of association between transition times. In this section, Markov renewal assumption will be further relaxed to allow the transition intensities to depend on the sojourn time of earlier states. Since it is an extension of Markov renewal model, this model will be defined as extended Markov renewal Model. Similar to Markov renewal model, the “clock-reset” approach will be used for the definition of time t in the hazard function $\alpha(t)$.

4.1 Estimation of Parameters

Similar to the prognostic covariates, effects of sojourn time in earlier states can be estimated by employing Cox proportional hazard model. For a patient with associated prognostic covariate vector $Z$ and time vector $J$, the transition hazard $\alpha_{gh}(t)$ for transition $g \rightarrow h$ is given by

$$\alpha_{gh}(t) = \alpha_{gh,0}(t) \exp(\beta^T Z_{gh} + \gamma^T J_{gh})$$

where $\alpha_{gh,0}(t)$ is the baseline hazard of transition $g \rightarrow h$, $\beta$ and $\gamma$ are vectors of regression coefficient respectively corresponding to $Z_{gh}$ and $J_{gh}$; and $J_{gh}$ is the vector of sojourn times specific to transition $g \rightarrow h$. Note that $J_{gh}$ only contains sojourn time of states no later than the $g$th state.

Similarly to the procedure introduced in Section 1.4, estimates $\hat{\beta}$, $\hat{\gamma}$ and the cumulative baseline hazards $\hat{A}_{gh,0}(t)$ can be obtained together by maximizing the generalized Cox partial likelihood

$$L(\beta, \gamma) = \prod_{\text{transition } g \rightarrow h} \prod_{i=1}^{n} \frac{\exp(\beta^T Z_{gh,i} + \gamma^T J_{gh,i})}{\sum_{j \in R_g(t_{gh,i})} \exp(\beta^T Z_{gh,j} + \gamma^T J_{gh,j})}$$

where $t_{gh,i}$ is the failure or censoring time of individual $i$ for transition $g \rightarrow h$, $d_{gh,i} = 1$ if individual $i$ has an event for transition $g \rightarrow h$, 0 otherwise, and $R_g(t_{gh,i})$ is the risk set of state $g$ at time $t$, i.e. the set of individuals who are in state $g$ at time $t$ ($t$ being here the time since entry in state $g$). The estimate of the cumulative baseline hazard of transition $g \rightarrow h$ is the Nelson-Aalen estimate:

$$\hat{A}_{gh,0}(t) = \sum_{i=1}^{n} \frac{d_{gh,i}}{\sum_{j \in R_g(t_{gh,i})} \exp(\beta^T Z_{gh,j} + \gamma^T J_{gh,j})}.$$
4.2 Prediction Formulas

The general problem is to estimate the conditional probabilities of some clinical future events, given the patient’s history and a set of values for prognostic factors $Z$. The estimate of these probabilities are based on the results obtained from the Cox model on the transition hazard between states.

It is not possible to write down explicitly the transition probability for a general non-Markov multi-state model. However in case of an illness death model where only three states are present (see Figure 3), it is possible to derive the prediction probability.

In this illness-death model, local recurrence, distant metastasis and the joint of both are taken together as one intermediate event, in short termed “Recurrence”. After surgery, a patient may die before or after experiencing tumor recurrence. The three possible states “Surgery”, “Recurrence” and “Death” are respectively numbered by 1, 2 and 3. In this model there are three possible paths that a patient may follow after surgery: from surgery to recurrence ($1 \rightarrow 2$); direct transition from surgery to death ($1 \rightarrow 3$); from surgery to recurrence to death ($1 \rightarrow 2 \rightarrow 3$).

The probabilities can be expressed in terms of the hazard rate for each transition.

For a patient without recurrence to $s$ years post-surgery, the probability that the patient remains in state 1 in the time interval $(s, t]$ is given by

$$P_{11}(s, t|Z) = \exp\left(- \int_s^t (\alpha_{12}(u|Z) + \alpha_{13}(u|Z))du\right), \quad (3)$$

where $\alpha_{12}(u|Z)$ and $\alpha_{13}(u|Z)$ are the transition hazards respectively corresponding to transition $1 \rightarrow 2$ and $1 \rightarrow 3$ given the patient’s covariates, $Z$ is the vector of covariates.

The conditional probability of being in state 2 at time $t$ given an individual is in state 1 at time $s$ is as follow

$$P_{12}(s, t|Z) = \int_s^t \alpha_{12}(u|Z)P_{11}(s, u^-|Z)P_{22}^u(u, t|Z)du. \quad (4)$$
The probability of remaining in state 2 in the time interval \((s,t]\) can be computed as
\[
P_{22}^r(s,t|Z) = \exp\left(-\int_s^t \alpha_{23,r}(u - r|Z) du\right),
\]
where \(r\) is the entrance time of state 2 \((r \leq s)\) and
\[
\alpha_{23,r}(u|Z) = \alpha_{23,0}(u) \exp(\beta^T Z + \gamma_2 r).
\]
The probability of death is given by
\[
P_{23}^r(s,t|Z) = 1 - P_{22}^r(s,t|Z).
\]
There are two possible paths going from state 1 to state 3. To distinguish them, we denote the probability of direct transition from surgery to death as \(P_{13}(s,t)\),
\[
P_{13}(s,t|Z) = \int_s^t \alpha_{13}(u|Z) P_{11}(s,u^-|Z) du.
\]
The probability that tumor recurrence and later on death occur during time interval \((s,t]\) is given by
\[
P_{13}^2(s,t|Z) = \int_s^t \alpha_{12}(u|Z) P_{11}(s,u^-|Z) P_{23}^u(u,t|Z) du.
\]
Note that for the transition probability \(P_{23}^u(u,t|Z)\) is as in (6).

### 4.3 Standard Error of Prediction

In Fiocco et al. [12], a simulation-based approach is proposed to obtain confidence intervals (CIs) for the estimated prediction probabilities. This method can generate paths through the multi-state model, and build bootstrap data sets that in turn can be used to obtained CIs for the estimate of interest.

In this section, a brief description how to generate path through a multi-state model based on given cumulative hazard functions specified for each of the direct transitions and how to estimate the standard errors of predicted probabilities by applying bootstrap resampling method will be given on. The basic idea is inspired by Dabrowsa [11] where a multi-states model is seen as made of several competing risk blocks linked together and to simulate transition times and states for each such block in the multi-state model.

Generating simulated path for a specific patient can be done as follow
Algorithm 4.3.1

Repeat, for \( m = 1, \ldots, M \),

1. Let \( \mathcal{J} \) be the set of states that can be reached from state \( g \). If \( \mathcal{J} = \emptyset \), stop. Otherwise, let, for \( h \in \mathcal{J} \), \( A_{gh}(t) \) be the cumulative hazard function for transition \( g \rightarrow h \).

2. Compute \( A_g(t) = \sum_{h \in \mathcal{J}} A_{gh}(t) \).

3. Sample \( t^*(> T_g) \) from \( A_g(t) - A_g(t_g) \). If \( A_g(\infty) \) is finite, \( t^* = \infty \) may be sampled with positive probability.

4. Stop if \( t^* = \infty \). Otherwise, select state \( h \) as the next state with probability \( \frac{dA_{gh}(t^*)}{dA_g(t^*)} \).

5. Set \( g = h \) and \( T_g = t^* \).

6. Repeat 1-5 until no further state can be reached or \( t^* = \infty \) is sampled.

Save the simulated path as \( P_m \).

This process can be also used to estimate prediction probabilities for an extended Markov renewal multi-state model, but it is less efficient than computations based on (3-8), since the simulation number \( M \) need to be large enough to obtain precise estimation. Another important use of this process is to obtained CIs of prediction possibilities, by combining with bootstrap resampling.

Let \( B \) be the number of bootstrap samples. The algorithm is described as follow.

Algorithm 4.3.2

Repeat, for \( b = 1, \ldots, B \),

1. Create a resampled data set \( X_b^* \) with replacement.

2. From \( X_b^* \), estimate the regression coefficients \( \hat{\beta}_b^* \), \( \hat{\gamma}_b^* \), and the baseline hazard functions \( \hat{A}_{gh,0}^*(t) \) for all \( g \rightarrow h \) transitions in the model.

3. Calculate the patient specific hazard function \( \hat{A}_{gh}^*(t) = \hat{A}_{gh,0}^*(t) \exp(\hat{\beta}_b^* Z_{gh} + \hat{\gamma}_b^* J_{gh}) \).

4. Simulate \( M \) paths through the multi-state model from \( \hat{A}_{gh}^* \) based on Algorithm 4.3.1, and estimate the prediction probabilities \( \hat{P}_M^* \). Set \( \hat{P}_{M,b}^* \) equal to \( \hat{P}_M^* \).

Once all probabilities \( \hat{P}_{M,b}^* \), \( b = 1, \ldots, B \) have been simulated, the 95% confidence interval for \( \hat{P}_M^* \) can be obtained by compute 2.5%-quantile and 97.5%-quantile of the vector \( \{ \hat{P}_{M,b}^* | b = 1, \ldots, B \} \). Confidence interval instead of standard error is used because probabilities are bounded by [0,1].

It is important to note that the use of \( M \) can be (much) smaller than \( M \). Because the computation of standard errors for \( \hat{P}_M^* \) involves two nested sets of simulations, the creation of bootstrap data sets and the subsequent simulations within each data set to obtain the bootstrap prediction probability \( \hat{P}_M^* \). (See Fiocco et al. [12]
An alternative to Algorithm 4.3.2 is to apply (3-8) instead of simulating paths to estimate prediction probabilities. After probabilities $\hat{P}_b^*$ are estimated for all resample data set $X_1^*, ..., X_b^*$, 95% CIs can be obtained in the same way as described above.

5 Application

In this section, the extended Markov renewal model will be employed to analyze breast cancer data from EORTC-trail. All analysis have been performed in R.

Table 1: Prognostic factors for all patients (n = 2795)

<table>
<thead>
<tr>
<th>Prognostic factor</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tumor size</strong></td>
<td></td>
</tr>
<tr>
<td>≤2 cm</td>
<td>823 (30)</td>
</tr>
<tr>
<td>2-5 cm</td>
<td>1759 (64)</td>
</tr>
<tr>
<td>&gt; 5 cm</td>
<td>166 (6)</td>
</tr>
<tr>
<td>Missing</td>
<td>47</td>
</tr>
<tr>
<td><strong>Nodal status</strong></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>1467 (53)</td>
</tr>
<tr>
<td>Positive</td>
<td>1327 (47)</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
</tr>
<tr>
<td><strong>Type of surgery</strong></td>
<td></td>
</tr>
<tr>
<td>Mastectomy RT</td>
<td>658 (24)</td>
</tr>
<tr>
<td>Mastectomy, no RT</td>
<td>577 (21)</td>
</tr>
<tr>
<td>Breast conserving</td>
<td>1560 (56)</td>
</tr>
<tr>
<td>Missing</td>
<td>16</td>
</tr>
<tr>
<td><strong>Perioperative chemotherapy</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1395 (50)</td>
</tr>
<tr>
<td>Yes</td>
<td>1398 (50)</td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
</tr>
<tr>
<td><strong>Adjuvant chemotherapy</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2227 (82)</td>
</tr>
<tr>
<td>Yes</td>
<td>502 (18)</td>
</tr>
<tr>
<td>Missing</td>
<td>66</td>
</tr>
<tr>
<td><strong>Age(years)</strong></td>
<td></td>
</tr>
<tr>
<td>≤50</td>
<td>1118 (40)</td>
</tr>
<tr>
<td>&gt; 50</td>
<td>1677 (60)</td>
</tr>
</tbody>
</table>

("RT" stands for radiation treatment)

5.1 Data Description

The dataset originates from a clinical trial in breast cancer, conducted by European Organization for Research and Treatment of Cancer (EORTC-trail 10854)
The aim of the trial was to study whether a short intensive course of pre-operative polychemotherapy yields better therapeutic result than surgery alone. The trial included women with early breast cancer, who underwent either radical mastectomy or breast conserving therapy before being randomized. The Trial consisted of 2795 patients, randomized to either chemotherapy or no chemotherapy. Details of the trial [13] and long-term results [14] can be found in [13-14]. Median follow-up was 10.8 years. The most important prognostic factors are shown in Table 1. Most of these factors contain a small number of missing values. Our analysis will be based on the patients with full information (n=2687, 96.1%). The illness-death model described in Section 2.2.2 (Figure 3) was applied to this data set. The number of patients to enter and to visit each state are shown in Table 2.

### Table 2: Number of patients to enter and visit the states.

<table>
<thead>
<tr>
<th>State</th>
<th>No. to enter</th>
<th>No. to visit</th>
<th>No. to visit</th>
<th>No. to visit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2687</td>
<td>-</td>
<td>-</td>
<td>1060(39.5%)</td>
</tr>
<tr>
<td>2</td>
<td>1060</td>
<td>-</td>
<td>-</td>
<td>645(60.8%)</td>
</tr>
<tr>
<td>3</td>
<td>84</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

5.2 Formats of Presenting Data

As a preliminary, two ways of representing the same data will be introduced in this section. The first of these is the one-row-per-subject format (the ‘wide’ format). Here below an example for the first three patients from the data set under study is shown:

<table>
<thead>
<tr>
<th>id</th>
<th>time_rec</th>
<th>status_rec</th>
<th>time_surv</th>
<th>status_surv</th>
<th>periop</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-189</td>
<td>2.007</td>
<td>1.000</td>
<td>3.978</td>
<td>no periop chemo</td>
</tr>
<tr>
<td>2</td>
<td>-188</td>
<td>13.380</td>
<td>0.000</td>
<td>13.380</td>
<td>periop chemo</td>
</tr>
<tr>
<td>3</td>
<td>-187</td>
<td>1.276</td>
<td>1.000</td>
<td>3.433</td>
<td>no periop chemo</td>
</tr>
</tbody>
</table>

The variables time_rec and time_surv are used to indicate occurrence (or censoring) time of tumor recurrence and death respectively. The variables status_rec and status_surv are used to indicate whether the occurrence of events has been
observed (1 observed event and 0 for censored). An alternative way of representing the same data is in the so-defined ‘long’ format. This format allows most flexibility for multi-state modeling. The same data in the ‘long’ format is as follow

<table>
<thead>
<tr>
<th>id</th>
<th>from</th>
<th>to</th>
<th>trans</th>
<th>Tstart</th>
<th>Tstop</th>
<th>time</th>
<th>status</th>
<th>periop.1</th>
<th>periop.2</th>
<th>periop.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-189</td>
<td>1</td>
<td>2</td>
<td>0.000</td>
<td>2.007</td>
<td>2.007</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>-189</td>
<td>1</td>
<td>3</td>
<td>0.000</td>
<td>2.007</td>
<td>2.007</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>-189</td>
<td>2</td>
<td>3</td>
<td>2.007</td>
<td>3.978</td>
<td>1.971</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>-188</td>
<td>1</td>
<td>2</td>
<td>0.000</td>
<td>13.380</td>
<td>13.380</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>-188</td>
<td>1</td>
<td>3</td>
<td>0.000</td>
<td>13.380</td>
<td>13.380</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>-187</td>
<td>1</td>
<td>2</td>
<td>0.000</td>
<td>1.276</td>
<td>1.276</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>-187</td>
<td>1</td>
<td>3</td>
<td>0.000</td>
<td>1.276</td>
<td>1.276</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>-187</td>
<td>2</td>
<td>3</td>
<td>1.276</td>
<td>3.433</td>
<td>2.157</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

For each individual a row is added for each possible transition at each state. i.e. For patient 189, who entered state 2 at $t = 2.007$, two transitions can occur from state 1: 1 → 2 and 1 → 3; it is possible to move only to state 3 from state 2: 2 → 3. For patient 188, who never entered state 2 during the follow-up, only two transitions from state 1 are possible. There are three rows corresponding to patient 189 but only two rows to patient 188. The variable `trans` was added with a unique value to label all possible transitions transitions (1 for transition 1 → 2, 2 for transition 1 → 3, and 3 for transition 2 → 3). The variables `from` and `to` are used to indicate the starting state and ending state for each transition respectively. The variables `Tstart` and `Tstop` indicate the entering time and leaving time for each transition. The variable `time` is the sojourn time spent in each state, and `status` is used to indicate whether the event was observed or censored. Extra dummy variables (`periop.1`, `periop.2`, `periop.3`) are transition-specific covariate (see Section 2.4). They have value 0 except for the patient’s condition for the transition that they correspond to. For instance, patients who did not receive perioperative chemotherapy are treated as reference group and dummies are coded as 0 for this treatment group. For patient 188 who received perioperative chemotherapy, `periop.1 = 1` for transition 1 but 0 for the others, `periop.2 = 1` for transition 2 but 0 for others.

**5.3 Data Analysis**

Cox regression model was fitted to the data by including all prognostic covariates and the sojourn time covariate. The results are summarized in Table 3.

Positive nodal status significantly increases transition rates for all transitions (0.443(0.075) for transition 1 → 2, 0.801(0.253) for transition 1 → 3, 0.750(0.095) for transition 2 → 3). Large tumor size has similar effects, but it is not significant.
<table>
<thead>
<tr>
<th>Covariate</th>
<th>1 → 2 Coef(SE) P-value</th>
<th>1 → 3 Coef(SE) P-value</th>
<th>2 → 3 Coef(SE) P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤2 cm</td>
<td>0.296(0.075) &lt;0.0001</td>
<td>-0.035(0.260) 0.89</td>
<td>0.122(0.106) 0.25</td>
</tr>
<tr>
<td>2-5 cm</td>
<td>0.771(0.132) &lt;0.0001</td>
<td>0.733(0.420) 0.081</td>
<td>0.305(0.167) 0.067</td>
</tr>
<tr>
<td>&gt; 5 cm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nodal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>0.443(0.075) &lt;0.0001</td>
<td>0.801(0.253) 0.002</td>
<td>0.750(0.095) &lt;0.0001</td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mast, RT</td>
<td>0.006(0.097) 0.95</td>
<td>1.021(0.312) 0.001</td>
<td>0.159(0.118) 0.18</td>
</tr>
<tr>
<td>Mast, no RT</td>
<td>-0.012(0.081) 0.88</td>
<td>0.089(0.312) 0.77</td>
<td>-0.123(0.100) 0.22</td>
</tr>
<tr>
<td>BCT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perioperative</td>
<td>No</td>
<td>-0.145(0.0615) 0.019</td>
<td>-0.131(0.219) 0.55</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjuvant chemotherapy</td>
<td>No</td>
<td>-0.295(0.103) 0.004</td>
<td>0.436(0.397) 0.27</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age(years)</td>
<td>≤50</td>
<td>-0.159(0.076) 0.037</td>
<td>0.617(0.305) 0.043</td>
</tr>
<tr>
<td></td>
<td>&gt; 50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of recurrence</td>
<td></td>
<td>-0.153(0.0186) &lt;0.0001</td>
<td></td>
</tr>
</tbody>
</table>
for transitions to death. Young age is a well know risk factor for recurrence and this can be seen from Table 3 (-0.159(0.076)), as well as the reverse effect for transition to death (0.617(0.305)). Perioperative chemotherapy and Adjuvant chemotherapy significantly decrease the transition rate for transition from surgery to tumor recurrence. Mastectomy without radiation treatment increase the transition rate from surgery to death comparing to Mastectomy with radiation treatment. The significant estimated time coefficient (-0.153(0.0186)) indicates violation of Markov property: the time at which tumor recurrence occurred has a significant effect on transition from recurrence to death. Early recurrence increases the risk of death.

The estimated baseline survival curves for all three transitions (values of co-

Figure 4: Baseline survival curves for all transitions
6 Software for Prediction in Extended Markov Renewal Model

6.1 Implementation for Prediction Formulas

To implement prediction for the extended Markov renewal model in R, new functions `getlevel()`, `dataprep()`, `predict()`, `plot.predict()` and `prob.predict()` were written for estimating and plotting prediction results. In this section, brief introduction and examples will be given. The detailed codes can be found in Appendix A.

Suppose we are interested on the future disease development of a patient A who has the following prognostic covariates: tumor size >5 cm, positive lymph node status, mastectomy plus radiotherapy, no perioperative chemotherapy, no adjuvant chemotherapy, age ≤ 50 years. The input for the patient A’s data in R (which is usually in ‘wide’ format) is as following

```r
> print(y)
periop surgery tusi nodal adjchem age50
1 no periop chemo mastectomy with RT >5 cm node positive no adj chemo <=50
```

The first step is to prepare the data into the long format with transition-specific covariates (see Section 1.2). The function `dataprep()` can rearrange the data from wide format into ‘long’ format, by specifying the data to be rearranged, names of the covariates, the fitted multi-state model, and the levels of categorical variables. For categorical variables the function `getlevel()` will give levels for all variables of interest at once. An example of how the two functions work is shown below. Use the function `getlevel()` to obtain the variable levels.

```r
> load("bc.long")
> covs<-c("periop", "surgery", "tusi", "nodal", "age50", "adjchem")
> (lvls<-getlevel(covs,bc.long))

$periop
[1] "no periop chemo" "periop chemo"

$surgery
[1] "mastectomy with RT" "mastectomy without RT" "breast conserving"

$tusi
[1] "<2 cm" "2-5 cm" ">5 cm"
```
The function `dataprep()` transforms the new data into the wide format with transition-specific covariates.

```r
> fit<-coxph(Surv(time,status)~-trans+strata(trans), bc.long[-c(1:3,6:15)])
> print(newdata<-dataprep(y,covs,fit,lvls))
```

The function `dataprep()` can also be used to prepare data of a single patient for the function `msfit()` in the `{mstate}` package [16], which is used to compute subject-specific or overall cumulative transition hazards for each of the possible transitions in Markov multi-state model.

After the data is reshaped into wide format the function `predict()`, which is based on the formulas given in Section 4.2, can be used to predict transition probabilities given the patient’s history at time point (t). If a patient A is alive without tumor recurrence at 2 years post-surgery, the predicted transition probabilities can be obtained by the following codes:
> results<-predict(fit,newdata,t=2,bt="Tstart")

The argument bt="Tstart" is to specify name of the estimated coefficient corresponding to time variable.
The returned object *results* contains two objects: a data frame called “probs” which contains time points and the corresponding estimated transition probabilities and a list called “hazards” which contains the estimated cumulative hazards. Examples of the estimated transition probabilities are shown below

<table>
<thead>
<tr>
<th>time</th>
<th>P11</th>
<th>P122</th>
<th>P123</th>
<th>P13</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>2.726899</td>
<td>0.7950662</td>
<td>0.1646767</td>
<td>0.03061893</td>
</tr>
<tr>
<td>[2,]</td>
<td>2.973306</td>
<td>0.7423555</td>
<td>0.1889760</td>
<td>0.05474760</td>
</tr>
<tr>
<td>[3,]</td>
<td>3.370294</td>
<td>0.6612630</td>
<td>0.2207091</td>
<td>0.10165261</td>
</tr>
<tr>
<td>[4,]</td>
<td>4.284736</td>
<td>0.5261654</td>
<td>0.2261588</td>
<td>0.22322631</td>
</tr>
<tr>
<td>[5,]</td>
<td>6.409309</td>
<td>0.3508181</td>
<td>0.1709543</td>
<td>0.44620436</td>
</tr>
</tbody>
</table>

P11, P122, P123, P13 respectively stand for probability of disease-free survival, probability of survival with tumor recurrence, probability of death after tumor recurrence and probability of death without tumor recurrence.

For a patient who has already experienced recurrence, the occurrence time of recurrence needs to be specified for the *Tstart* variable. The returned list of prediction probabilities only contains two columns: P22 and P23, which are respectively the probability of survival with recurrence and the probability of death. Function *predict()* can be used to predict transition probabilities for Markov renewal model, by specifying the time coefficient bt = NULL.

The function `plot.predict()` produces survival curves by using the output of `predict`. An example is shown in Figure 5. The plot shows the predicted transition probabilities for patient A given that no events have occurred during the first two years after surgery. The probabilities are stacked; the height of each band is the probability of the path. The paths are ordered from top to bottom according to increasing disease severity.

The function `prob.predict()` produces estimated transition intensities at certain time point $t$ by using output from the function `predict()`. Suppose that clinicians are interested in predicting the specific status for patient A at $t = 6$ years post-surgery. The transition intensities can be obtained as:

> `prob.predict(t=6,results$probs)`

```
P1    P2    P3
0.3807090 0.1763831 0.4420988
```

24
where P1, P2, P3 respectively stand for probability of disease-free survival, probability of survival with recurrence and probability of death, the subscript numbers indicating the state occupied at time $t$.

### 6.2 Implementation for Simulation Method

In [12] the functions `mssample()` and `msboot()` have been used to compute prediction probabilities and their confidence intervals and standard errors of the regression coefficients in multi-state reduced rank models. The same procedure will be employed to estimate the confidence intervals for prediction probabilities in our extended Markov renewal model. A brief description of the two functions will be given in this section. All details concerning the R code can be found in Appendix B.

The function `mssample()` implements Algorithm 4.3.1 to generate a specified
number of paths through the multi-state model and computes probabilities of states and paths given the starting state and time. As mentioned in Section 4.3, `mssample()` can also be used to obtain prediction of transition probabilities but it is less efficient than `predict()`. Figure 6 shows the prediction probabilities by respectively employing the two methods for a patient B who just underwent surgery and who has the same prognostic covariates as patient A. As expected, the prediction results are very close when the simulation number \(M\) is large. Compared to simulations by using function `mssample`, `predict()` can produce smoother curves and provide trajectory probabilities instead of transition intensities.

![Figure 6: Prediction probabilities by implementing formula (left) and simulation (right)](image)

The function `msboot()` samples randomly with replacement subjects from the original data set and can be used to estimate any vector-valued statistic in a multi-state model. It can be combined together with `mssample()` to implement the Algorithm 4.3.2. Here below it is shown how to obtain prediction probabilities and confident interval for patient B (tumor size >5 cm, positive lymph node status, mastectomy plus radiotherapy, no perioperative chemotherapy, no adjuvant chemotherapy, age ≤ 50 years) at \(t = 6\) years post-surgery by using the functions `predict()`, `mssample()`, and `msboot()`:

```r
> tmat <- trans.illdeath()#specify the transition matrix
```
> M=100  # simulation number
> tvec=6  # time point
> history<-list(state=1, time=0, tstates=c(0,0,0))  # specify the patient's history
> theta<-function(data){
+   dat<-data[, -c(1:3,6,9:15)]
+   fit<-coxph(Surv(time,status) ~ . - trans + strata(trans), dat)
+   Haz<-haz(fit, newdata)
+   bt<-summary(fit)$coef[,1]["Tstart"]
+   beta.state<-matrix(0,3,3)
+   beta.state[1,2]<-bt
+   res<-mssample(Haz, trans=tmat, clock="reset",
+                  history=history, beta.state=beta.state, tvec=tvec, M=M)
+   c(as.matrix(res[,-1]))
+ }
> res<-msboot(theta, bc.long, B=500, id="id")  # generate bootstrap samples
> predB<-predict(fit, newdata, t=0, bt="Tstart")
> estimate<-prob.predict(t=6, predB$probs)
> LCI<-apply(res, 1, function(x) quantile(x, probs=0.025))
> UCI<-apply(res, 1, function(x) quantile(x, probs=0.975))
> print(CI<-cbind(LCI, estimate, UCI), digits=2)

<table>
<thead>
<tr>
<th></th>
<th>LCI</th>
<th>estimate</th>
<th>UCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0.08</td>
<td>0.24</td>
<td>0.43</td>
</tr>
<tr>
<td>P2</td>
<td>0.04</td>
<td>0.12</td>
<td>0.21</td>
</tr>
<tr>
<td>P3</td>
<td>0.44</td>
<td>0.64</td>
<td>0.86</td>
</tr>
</tbody>
</table>
7 Results and Conclusions in the Extended Markov Renewal Model

Three fictitious patients are used to show predicted probabilities: patient A, B and C with a common set of prognostic covariate values (tumor size >5 cm, positive lymph node status, mastectomy plus radiotherapy, no perioperative chemotherapy, no adjuvant chemotherapy, age ≤ 50 years). For each patient, both extended Markov renewal model and Markov renewal model were employed to predict future trajectory probabilities.

Figure 7a and 7b show the predicted trajectory probabilities for patient A given that no events have occurred during the first two years after surgery respectively for extended Markov renewal model and Markov renewal model. Compared to the extended Markov renewal model, Markov renewal model yields to more pessimist prediction results for disease-free survival and survival with tumor recurrence. Markov renewal model ignores the association between time at which recurrence occurs and the transition rate from recurrence to death. As mentioned in Section 3.3, early recurrence increases the risk of death. Therefore patients who experience recurrence at later time have lower transition rate to death.

Figures 7c and 7d show the predicted trajectory probabilities for patient B who just underwent surgery respectively for extended Markov renewal model and Markov renewal model. Similarly to prediction results for patient A above, Markov renewal model shows more pessimist estimation for survival with and without recurrence comparing to the extended Markov renewal model. Figure 8 shows the predicted state probabilities with 95% confidence intervals by employing the extended Markov renewal model. The probability of being alive with recurrence initially increases but later on decreases as time $t$ increases. The transition probability from surgery to recurrence initially increases and later on decreases, whereas the transition probability from recurrence to death increase rapidly as time $t$ increases.

Figures 7e and 7f show the predicted trajectory probabilities for patient C who has experienced a tumor recurrence and no other event within $t = 5$ years after surgery respectively for extended Markov renewal model and Markov renewal model. The influence of time $r$ at which recurrence occurred is shown in Figure 7e, where the survival probabilities assuming recurrence occurred at $t = 0.5, 1, 2, 4$ years post-surgery are illustrated. Comparing to a patient who had recurrence half year after surgery a patient who had recurrence four years after surgery has higher survival probability.
(a) Prediction by extended Markov renewal model: (b) Prediction by Markov renewal model: patient A

(c) Prediction by extended Markov renewal model: (d) Prediction by Markov renewal model: patient B

(e) Prediction by extended Markov renewal model: (f) Prediction by Markov renewal model: patient C

Figure 7: Predicted probability if future trajectories for patient A, B and C
(a) Probability of disease-free survival  

(b) Probability of survival with recurrence  

(c) Probability of death  

Figure 8: Predicted state probability with 95% CI for patient B
8 Frailty Model

In the previous sections, we have introduced an extension of Markov renewal model to deal with possible association between transitions. An alternative approach to model the association between transitions is throughout frailty models.

A frailty model is a model with random effects (frailty) which acts multiplicatively on the hazard. Frailty can be interpreted as an unobserved term affecting the transition speed of an individual or a group or cluster of individuals. Individuals with higher value of frailty have larger hazards and their corresponding risk of death is higher. Frailties can be used to explain effects of unobserved or unobservable heterogeneity caused by different sources like clustered data or deviation from proportional hazards assumption.

In multi-state model frailties can be used for different purposes. In an early paper of Aalen [21] frailty was applied to a time-homogeneous Markov model as a shared random term which affects the speed of the Markov process across all transitions. Bhattacharya and Klein [22] used correlated gamma frailties to model the associations between transition intensities for a non-homogeneous Markov Model. A paper of Yen et al. [23] applied frailty to account for the different underlying propensity for progression of premalignant lesions. Putter et al. [25] discussed the role of frailty in competing risk and in sequence of events by employing two frailty distributions: gamma distributed frailties and two-point mixture frailties. In this dissertation we will combine the two-point mixture frailties in [25] with hidden Markov model for a forward sequential process.

8.1 Frailties in simple survival models

Let $T$ denote the survival time and suppose that conditional on a frailty $Z$ the hazard of dying is given as follow

$$\lambda(t|Z) = Z\lambda(t).$$

(9)

The latent frailty is assumed to act in a multiplicative way on the hazard; $\lambda(t)$ is the conditional hazard given $Z = 1$.

At population level the frailty induces selection because individuals with high frailty die first and individuals who are still alive at time $t$ ($t > 0$) have thus lower average frailty than the population average at the start. The marginal hazard is given by

$$\lambda^*(t) = \lambda(t)E(Z|T > t).$$

(10)

A very convenient tool for deriving properties of the population is the Laplace transform of the frailty distribution [24] which is defined as

$$\mathcal{L}(c) = E\exp(-cZ).$$

(11)
The marginal survival function $S^*(t)$ is defined as

$$ S^*(t) = P(T > t) = E \exp(-Z \Lambda(t)) = \mathcal{L}(\Lambda(t)), \quad (12) $$

where $\Lambda(t)$ is the conditional cumulative hazard given $Z = 1$.

By the relation $\lambda^*(t) = -d \log S^*(t)/dt$ the marginal hazard $\lambda^*(t)$ can be expressed in terms of the Laplace transform as

$$ \lambda^*(t) = \lambda(t) \cdot \frac{-\mathcal{L}'(\lambda(t))}{\mathcal{L}(\lambda(t))}. \quad (13) $$

### 8.2 Two-point Mixture Frailty Distribution

Many distributions can be chosen for frailty. The most commonly used frailty distribution is the gamma distribution for its convenience in mathematical computation. In this dissertation we will focus on the two-point mixture frailty with distribution $P(Z = 1) = 1 - \pi$ and $P(Z = \theta) = \pi$, which is less often used but also convenient from computational and analytical perspective. For identifiability reasons, it is assumed that $\theta > 1$. Its expectation is $\pi(\theta - 1) + 1$, and the variance is $\pi(1 - \pi)(\theta - 1)^2$. The Laplace transform is given by

$$ \mathcal{L}(c) = E \exp(-cZ) = (1 - \pi)e^{-c} + \pi e^{-\theta c}. \quad (14) $$

From (12) the marginal survival function is

$$ S^*(t) = \mathcal{L}(\Lambda(t)) = (1 - \pi)e^{-\Lambda(t)} + \pi e^{-\theta \Lambda(t)} \quad (15) $$

From (13) the marginal hazard rate is

$$ \lambda^*(t) = \lambda(t) \cdot \frac{1 - \pi + \theta \pi e^{-(\theta - 1)\Lambda(t)}}{1 - \pi + \pi e^{-\theta \Lambda(t)}}. \quad (16) $$

The fraction of alive population changes over time as a result of selection. The population fraction for patients that survive to time $t$ $P(Z = \theta|T > t)$ is given as

$$ P(Z = \theta|T > t) = \frac{P(Z = \theta, T > t)}{P(T > t)} = \frac{\pi e^{-\theta \Lambda(t)}}{(1 - \pi)e^{-\Lambda(t)} + \pi e^{-\theta \Lambda(t)}}. \quad (17) $$

32
9 Hidden Markov Model

A hidden Markov model (HMM) is a statistical model in which the underlying stochastic process is assumed to be Markovian but not observable (“hidden”). The underlying stochastic process can produce another set of observed variables, which could be either discrete or continuous. To define a hidden Markov model, the definitions and formalism in Rabiner and Juang [26] will be used.

Denote by $Q = \{Q_1, Q_2, ..., Q_N\}$ the hidden states at a sequence of time points $t_1 < t_2 < ... < t_N$. Let $O = \{O_1, O_2, ..., O_N\}$ be the observation sequence. Under an HMM two independence assumptions are made about the hidden and observed variables: (i) given the $k^{th}$ hidden variable the $(k + 1)^{th}$ hidden variable is independent of all previous variables, or

$$P(Q_{k+1}|Q_k, O_k, ..., Q_1, O_1) = P(Q_{k+1}|Q_k); \quad (18)$$

(ii) given the $k^{th}$ hidden variable the $k^{th}$ observation is independent of other variables or

$$P(O_k|Q_k, Q_{k-1}, O_{k-1}, ..., Q_1, O_1) = P(O_k|Q_k). \quad (19)$$

Based on assumption (i) the transition probability matrix can be defined as

$$A = \{\alpha_{gh}\}, \alpha_{gh} = P(Q_{k+1} = h|Q_k = g). \quad (20)$$

Based on assumption (ii)

$$B = \{b_g(k)\}, b_g(k) = P(O_k|Q_k = g). \quad (21)$$

where $b_g(k)$ is the probability of observation $O_k$ at time point $t_k$ in state $g$.

The initial state distribution at time point $t_1$ is defined as

$$\pi = \{\pi_g\}, \pi_g = P(Q_1 = g). \quad (22)$$

The complete set of $(A, B, \pi)$ can be used to specify an HMM. In application of HMM, there are three key problems of interest [26]:

1. Compute the probability of observation sequence given the HMM model $(A, B, \pi)$ and the observation sequence $O = O_1, ..., O_N$.
2. Find the best hidden states sequence $Q_1, ..., Q_N$ for a given observation sequence $O = O_1, ..., O_N$.
3. Find the optimal model parameters $(A, B, \pi)$ to maximize $P(O|A, B, \pi)$.  

33
10 Thesis Contribution: Hidden Markov Two Point Frailty Model

The idea behind two-point mixture frailty model assumes that there is a latent class of individuals who experienced the transition at a normal speed, as well as a class of individuals who experienced the transition at a higher speed. The first class can be referred as normal class or “class 0”, while the second class can be referred as frail class or “class 1”. The latent class to which an individual belongs cannot be observed but could influence the hazard rate and thereby be estimated from the occurrence of events and corresponding sojourn times. This plays a role similar to hidden states in an hidden Markov model. The difference between the two approaches is that in a two-point frailty model an individual belongs to one latent class across all transitions.

By extending the two-point frailty model we propose a hidden Markov two-point mixture model which allows latent frailty of patients varies for different transitions. The new model contains two hierarchies: the first part concerns the variation of frailties across different transitions and the second part concerns the hazard and survival function for given frailty in each transition.

In this section the hidden Markov two-point mixture model will be illustrated by applying it to a forward sequential process in breast cancer trial (see Figure 9).

![Figure 9: A multi-state model in sequence for a breast cancer trial.](image)

10.1 Definitions and Notations

In the multi-state state model shown in Figure 9 there are three transitions: from surgery to local recurrence (transition 1), from local recurrence to distant metastasis (transition 2), from distant metastasis to death (transition 3).

Let $t_{ij}$ and $\delta_{ij}$ denote time and status for individual $i$ and transition $j$ respectively with $i = 1, ..., n$ and $j = 1, 2, 3$. The indicator function $\delta_{ij}$ is defined as

$$\delta_{ij} = \begin{cases} 1, & \text{if transition } j \text{ occurs for patient } i \\ 0, & \text{otherwise} \end{cases}$$

It is obvious that transition 2 can be observed only after transition 1 has occurred ($\delta_{i1} = 1$). The same holds for transition 3 ($\delta_{i1} = 1$, $\delta_{i2} = 1$).
Denote the frailty class of individual $i$ corresponding to the three transitions by $G_i = \{g_{i1}, g_{i2}, g_{i3}\}$, where $g_{ij} = (0, 1)$ and $j = 1, 2, 3$. The interpretation of the frailty variables is that patients with frailty $g_j = 1$ ("class 1") travel through the $j^{th}$ transition at a faster speed comparing to patients with frailty $g_j = 0$ ("class 0"). Let $\theta_j = \exp(\gamma_j)$ denote the hazard ratio of the $j^{th}$ transition of class 1 with respect to class 0. For identifiability reasons it is assumed that $\theta_j > 1$ for $j = 1, 2, 3$.

The structure of the complete model is shown in Figure 10. At individual level the frailties $g_1, g_2, g_3$ are unobservable hidden states. The observations $t_j$ and status indicators $\delta_j$ depend on frailties through hazards rate $\lambda_j(t|g_1) = \lambda_{j0}(t) \exp(g_j\gamma_j)$ where $\lambda_{j0}(t)$ is the baseline hazard corresponding to the $j^{th}$ transition.

**Figure 10:** Hidden Markov: extension for the two-points frailties

### 10.2 Hidden Markov: extension for the two-points frailties

All notations and computations in this section are at individual level and the subscript $i$ is omitted for convenience. Under an HMM

$$P(g_2|g_1, t_1, \delta_1) = P(g_2|g_1)$$

and

$$P(g_3|g_1, t_1, \delta_1, g_2, t_2) = P(g_3|g_2).$$

Define $\pi = P(g_1 = 1)$ as the initiate state distribution. Denote the frailty transition probabilities as follow

$p_{01} = P(g_2 = 1|g_1 = 0), p_{10} = P(g_2 = 0|g_1 = 1)$

$p_{01} = P(g_3 = 1|g_2 = 0), p_{10} = P(g_3 = 0|g_2 = 1).$

For identifiability reason, it is assumed that

$p_{01} = p_{01} = p_{01}; p_{10} = p_{10} = p_{10}$

35
The joint probability $P(g_1, g_2, g_3)$ can be expressed as the product of initiate state probability and frailty transition probabilities. Table 4 lists all possible frailty sequences and corresponding probabilities.

<table>
<thead>
<tr>
<th>$g_1$</th>
<th>$g_2$</th>
<th>$g_3$</th>
<th>$P(g_1, g_2, g_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$(1 - \pi)(1 - p_{01})^2$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>$(1 - \pi)(1 - p_{01})p_{10}$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>$(1 - \pi)p_{01}p_{10}$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>$(1 - \pi)p_{01}(1 - p_{10})$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>$\pi p_{10}(1 - p_{01})$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>$\pi p_{10}p_{01}$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>$\pi (1 - p_{10})p_{10}$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>$\pi (1 - p_{10})^2$</td>
</tr>
</tbody>
</table>

Table 4: Possible latent classes and corresponding joint probabilities

More assumptions can be made to simplify the joint probabilities or for application purpose. For instance if it is assumed that $\pi = p_{01} = 1 - P(g_1 = g_2)$ then the resulting model can presents the correlation between transition 1 and transition 2. If $P(g_1 = g_2) = 0.5$ the two transitions have correlation equal to 0; if $P(g_1 = g_2) > 0.5$ the two transitions are positively correlated; if $P(g_1 = g_2) < 0.5$ the two transitions are negatively correlated; $P(g_1 = g_2) = 1$ and 0 indicates perfect positive correlation and perfect negative correlation respectively.

In case that $P(g_1 = g_2) = P(g_2 = g_3) = 1$ the resulting model will be equivalent to a two-point mixture model with three frailties.
11 Work in Progress: EM Algorithm

The complete set of parameters \((\gamma_1, \gamma_2, \gamma_3, \pi, p_{10}, p_{01})\) can be used to specify an
Hidden Markov two-point frailty model. To estimate these parameters, the Expectation - Maximization (EM) algorithm [27] is very attractive due to the complete data likelihood structure. In this section, an EM algorithm for the hidden Markov two-point frailty model will be briefly introduced. Further developments for implementation of the EM algorithm are still in progress.
Denote by \(l_{i;g_1, g_2, g_3}\) the likelihood for individual \(i\) conditional on the latent class \(G_i\) defined before. The transition rates for individual \(i\) are given by the following proportional hazards equations

\[
\lambda_{j;G_i} = \lambda_j(t_{ij}) \cdot \gamma_j g_{ij}
\]

where \(j = 1, 2, 3\).
For an individual \(i\) the corresponding likelihood will be

\[
l_{i;g_1, g_2, g_3} = \lambda_{1;G_i}(t_{i1}) \delta_{i1} \exp(-\Lambda_{1;G_i}(t_{i1})) \times \lambda_{2;G_i}(t_{i2}) \delta_{i2} \exp(-(\Lambda_{2;G_i}(t_{i2}) - \Lambda_{2;G_i}(t_{i1}))) \times \lambda_{3;G_i}(t_{i3}) \delta_{i3} \exp(-(\Lambda_{3;G_i}(t_{i3}) - \Lambda_{3;G_i}(t_{i1}) - \Lambda_{3;G_i}(t_{i2})))^{\delta_{i2} \delta_{i1}}. \tag{23}
\]

Based on (23) for an individual \(i\) who never experience transition 1 (\(\delta_{i1} = 0\), the conditional likelihood is given by

\[
l_{i;g_1, g_2, g_3} = \exp(-\Lambda_{1;G_i}(t_{i1})).
\]

Similarly, for an individual \(i\) who has experienced transition 1 but never experience transition 2 (\(\delta_{i1} = 1, \delta_{i2} = 0\), the conditional likelihood is given by

\[
l_{i;g_1, g_2, g_3} = \lambda_{1;G_i}(t_{i1}) \exp(-\Lambda_{1;G_i}(t_{i1})) \exp(-(\Lambda_{2;G_i}(t_{i2}) - \Lambda_{2;G_i}(t_{i1})));
\]

and for individual who has experience transition 1 and 2 but not transition 3, the conditional likelihood is given by

\[
l_{i;g_1, g_2, g_3} = \lambda_{1;G_i}(t_{i1}) \exp(-\Lambda_{1;G_i}(t_{i1})) \times \lambda_{2;G_i}(t_{i2}) \exp(-(\Lambda_{2;G_i}(t_{i2}) - \Lambda_{2;G_i}(t_{i1}))) \times \exp(-(\Lambda_{3;G_i}(t_{i3}) - \Lambda_{3;G_i}(t_{i2}) - \Lambda_{3;G_i}(t_{i1}))). \tag{24}
\]

For all possible latent classes, we multiply the (cumulative) hazards with the appropriate hazard ratios. The complete log-likelihood is as follow

\[
\mathcal{L}(\gamma_1, \gamma_2, \gamma_3, \pi, p_{10}, p_{01}) = \sum_{i=1}^{n} \log(P(G = 0, 0, 0)l_{i;0,0,0} + P(G = 0, 0, 1)l_{i;0,0,1} + P(G = 0, 1, 0)l_{i;0,1,0} + P(G = 0, 1, 1)l_{i;0,1,1} + P(G = 1, 0, 0)l_{i;1,0,0} + P(G = 1, 0, 1)l_{i;1,0,1} + P(G = 1, 1, 0)l_{i;1,1,0} + P(G = 1, 1, 1)l_{i;1,1,1}).
\]

37
EM algorithm

(i) **E-step** Start with the log-likelihood contribution for individual $i$. The conditional expectation of $\hat{\pi}_i$, given the data of individual $i$ can be calculated by

$$\hat{\pi}_i = \frac{P(G = 1, 0, 0)l_{i,1,0,0} + P(G = 1, 0, 1)l_{i,1,0,1}}{\sum P(G)l_{i,G}} + \frac{P(G = 1, 1, 0)l_{i,1,1,0} + P(G = 1, 1, 1)l_{i,1,1,1}}{\sum P(G)l_{i,G}},$$

where $P(G)$ can be computed according to Table 4 and where $l_{i,G}$ can be computed based on (24).

Similarly the conditional expectation of $\hat{p}_{i:01}$ given the data corresponding to individual $i$ can be calculated by

$$\hat{p}_{i:01} = \frac{P(G = 0, 1, 0)l_{i,0,1,0} + P(G = 0, 1, 1)l_{i,0,1,1}}{\sum P(G)l_{i,G}} + \frac{P(G = 1, 0, 1)l_{i,0,0,1} + P(G = 0, 0, 1)l_{i,0,0,1}}{\sum P(G)l_{i,G}}.$$

The conditional expectation of $\hat{p}_{i:10}$ is given as

$$\hat{p}_{i:10} = \frac{P(G = 0, 1, 0)l_{i,0,1,0} + P(G = 1, 1, 0)l_{i,1,1,0}}{\sum P(G)l_{i,G}} + \frac{P(G = 1, 0, 1)l_{i,0,0,1} + P(G = 1, 0, 0)l_{i,1,0,0}}{\sum P(G)l_{i,G}}.$$

(ii) **M-step** For the whole data set, calculate

$$\hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} \hat{\pi}_i, \quad \hat{p}_{01} = \frac{1}{n} \sum_{i=1}^{n} \hat{p}_{i:01}, \quad \text{and} \quad \hat{p}_{10} = \frac{1}{n} \sum_{i=1}^{n} \hat{p}_{i:10}.$$ 

The hazard ratios $\gamma_1$, $\gamma_2$ and $\gamma_3$ can be obtained separately by weighted Cox regressions for transition 1, 2, and 3 separately. The weights are employed here because the latent classes are not observable and only their distribution can be estimated. Let $w_j$ denote the vector of weights corresponding to transition $j$. The weights can be computed as
\[ w_{ij} = \frac{\sum_{G: g_j = 1} P(G) l_{i,G}}{\sum P(G) l_{i,G}} \]

(iii) **Repeat** Iteration stops when the log-likelihood \( \mathcal{L}(\gamma_1, \gamma_2, \gamma_3, \pi, p_{10}, p_{01}) \) for subsequent iterations differs by less than a pre-specified small value.
12 Discussion

The focus of this thesis is to address the violation of the Markov or Markov renewal assumption in applications of multi-state models. Two statistical methods were proposed. The first method was to include transition times as extra covariates, and the second was to introduce frailties under a Hidden Markov assumption.

By employing a Cox regression model the influence of previous sojourn times on later transitions can be estimated together with prognostic covariates effects. Explicit analytical expression for the trajectory probabilities was developed for the extended Markov renewal model. The parameter estimates and corresponding baseline hazards can be used to obtain prediction probabilities of future events. Confidence interval of the predicted probabilities can be obtained by simulations. The breast cancer data set was analyzed by employing an extended Markov renewal illness-death model (surgery, tumor recurrence and death). Prediction results between the extended Markov renewal model and classic Markov renewal model have been compared.

The merit of this extended model is the ability to deal with association between transition times. Illness-death model can be further extended to more complex model. However, the number of parameters to be estimated increases for more complex model, since the effects of sojourn time in each state on each later transition needs to be estimated in the extended Markov renewal model.

Another disadvantage of this approach is the computation-costly procedure for obtaining CIs by bootstrap sampling and simulation of trajectories through the multi-state model.

Mathematical expression to estimate the predicted probabilities has been developed in this dissertation. For Markov models, formulas based on Aalen and Johansen's estimator and their standard errors are available (see [21]). For Markov renewal models formulas for both non-parametric and semi-parametric estimators of the transition probabilities and corresponding standard error are recently available (see Spitoni at al. [20]). More research need to be done to compute standard error for the extended Markov model to provide a more efficient way of computing standard errors for prediction estimates.

It is important to investigate whether the transition time effects satisfies the Cox proportional assumption in applications of this extended Markov renewal model. The violation of Cox proportional assumption will lead to inconsistent estimate results. There are two possible solutions to this problem. The first solution is to stratify patients by their sojourn times in each transitions. Unfortunately this method can be applicable only for simple model with a few transitions.

Another solution is to introduce frailties which can deal with association between transitions as well as violation of Cox proportional hazard assumption. We proposed a combination of Hidden Markov model and two-point mixture frailty model.
One merit of the hidden Markov two-point frailty model is that it allows much flexibility when modeling the latent frailty classes for patients population. It can deal with the violation of the Cox proportional hazard assumption as well. One shortcoming of this method might be the difficulty of estimating parameters in an efficient way. We proposed an EM algorithm to estimate this model. The results seem promising but more research need to be done.
References


A Functions Defined for the Analysis

```r
getlevel <- function(covs, data) {
  # data is the original data, covs is the covs used to expand
  if (mean(covs %in% colnames(data)) != 1) {
    stop("Found: undefined variable")
  } else {
    temp <- as.list(data[, colnames(data) %in% covs])
    sapply(temp, levels)
  }
}

newprep <- function(y, covs, fit, lvls) {
  # covs is the vector covs when expand the data
  if (ncol(y) != length(covs)) {
    stop("The number of covariates do not match")
  } else {
    y <- y[, order(colnames(y))]
    covs <- sort(covs)
    if (sum(colnames(y) != covs) > 0) {
      stop("The covariate names do not match")
    } else {
      beta.names <- sort(names(summary(fit)$coef[, 1]))
      n.trans <- length(fit$xlevels[[1]])
      newdata <- matrix(0, n.trans * nrow(y), length(beta.names))
      colnames(newdata) <- beta.names
      for (i in 1:nrow(y)) {
        for (j in 1:ncol(y)) {
          temp <- paste(covs[j], sep = "") # extract name
          temp.1 <- lvls[[temp]]
          if (length(temp.1) > 0) {
            tempdummy <- which(temp.1 == y[i, j]) - 1 # character
            if (length(tempdummy) == 0) {
              stop("undefined variable")
            } else {
              inx <- grep(temp, beta.names)[1:n.trans] + (tempdummy - 1) * n.trans
              diag(newdata[(i - 1) * n.trans + (1:n.trans), inx]) <- y[i, j]
            }
          }
        }
      }
    }
  }
}
```
# numeric

```r
data.frame(trans=rep(1:n.trans,nrow(y)),newdata)
```

```r
S.t <- function(time, S) {
  t <- S[,1]
  s <- S[,2]
  n <- length(time)
  result <- numeric(n)
  for (i in 1:n) {
    if (time[i] < min(t)) result[i] <- 1
    else if (time[i] > max(t)) result[i] <- min(S)
    else result[i] <- tail(S[t<=time[i]], 1)
  }
  result
}
predict1 <- function(fit, newdata, t, bt, r) {
  newdata$trans <- NULL
  beta.all <- summary(fit)$coef[,1]
  if (is.null(bt)) {beta.t <- 0
    beta <- beta.all[order(names(beta.all))]
  } else if (sum(bt %in% names(beta.all)) == 0) stop("undefined time variable")
  else {beta.t <- beta.all[bt]
    beta <- beta.all[names(beta.all) != bt]
    beta <- beta[order(names(beta))]
    newdata <- newdata[, colnames(newdata) != bt]
  }
  if (mean(names(beta) == names(newdata)) != 1) stop("undefined variable")
  else {ZB <- as.matrix(newdata) %*% beta
```

46
fit.base <- survfit(fit)
strata <- fit.base$strata
inx <- c(cumsum(strata) - strata, sum(strata))

H.baseline <- sapply(1:3, function(x)
cbind(fit.base$time[(inx[x]+1):inx[x+1]], -log(fit.base$surv[(inx[x]+1):inx[x+1]]))
time <- H.baseline[[3]][,1]

H.Zr <- H.baseline[[3]][,2] * exp(ZB[3])*exp(r*beta.t)

S22 <- exp(-H.Zr)
P23 <- 1 - S22

result <- list(probs = cbind(time, S22, S23), hazards = H.Zr)
colnames(result[[1]]) <- c("time", "P22", "P23")
return(result)

predict0 <- function(fit, newdata, t, bt){
  newdata$trans <- NULL
  beta.all <- summary(fit)$coef[,1]
  if (is.null(bt)){beta.t <- 0
    beta <- beta.all[order(names(beta.all))]
  } else if (sum(bt %in% names(beta.all)) == 0) stop("undefined time variable")
  else {beta.t <- beta.all[bt]
    beta <- beta.all[names(beta.all) != bt]
    beta <- beta[order(names(beta))]}
    newdata <- newdata[, colnames(newdata) != bt]
  }

  if (mean(names(beta) == names(newdata)) != 1){stop("undefined variable")}
  else {ZB <- as.matrix(newdata) %*% beta}
```r
fit.base <- survfit(fit)
strata <- fit.base$strata
inx <- c(cumsum(strata) - strata, sum(strata))

# baseline
H.baseline <- sapply(1:3, function(x)
  cbind(fit.base$time[(inx[x]+1):inx[x+1]],
  -log(fit.base$surv[(inx[x]+1):inx[x+1]])))

# semi-Markov
H.Z <- sapply(1:3, function(x)
  cbind(H.baseline[[x]][,1],
  H.baseline[[x]][,2] * exp(ZB[x])))

S22 <- sapply(1:3, function(x) cbind(H.Z[[x]][,1],
  exp(-H.Z[[x]][,2])))

P23 <- 1 - S22

### P13 from state 1 to state 3

DFS

H1.Z <- H.Z[[1]][,2]
H2.Z <- H.Z[[2]][,2]
dH1.Z <- c(H1.Z[1], diff(H1.Z))
dH2.Z <- c(H2.Z[1], diff(H2.Z))
time1 <- H.Z[[1]][,1]
inx <- time1 >= t & (dH1.Z != 0 | dH2.Z != 0)
time.t <- time1[inx]
dH1.Zt <- dH1.Z[inx]
dH2.Zt <- dH2.Z[inx]
S11.Zt <- exp(-cumsum(dH1.Zt + dH2.Zt))

### P12 from state 1 to state 2 and stay

P122.Zt <- rep(NA, sum(inx))
for (i in 1:sum(inx)) {
  v <- time.t[1:i] - v
dtime <- time.t[i] - v
  S22.dtime.0 <- S.t(dtime, S22)
  S22.dtime.v <- S22.dtime.0 * exp(v * beta.t)
}
```
\[
P_{123} Z_t[i] \leftarrow \text{sum}(dH_1 Z_t[1:i] \ast S_{11} Z_t[1:i] \ast (1 - S_{22} \text{dtime} . v))
\]

result <- list(probs=cbind(time.t, S_{11} Z_t, P_{122} Z_t, P_{123} Z_t, P_{13} Z_t), hazards=H.Z)
colnames(result[[1]]) = c("time", "P11", "P122", "P123", "P13")
colnames(result$hazards[[1]]) = c("time", "trans1")
colnames(result$hazards[[2]]) = c("time", "trans2")
colnames(result$hazards[[3]]) = c("time", "trans3")
return(result)

predict <- function(fit, newdata, t, bt){
  if (newdata$Tstart[2] != 0)
    result <- predict1(fit, newdata, t, bt, r)
  else
    result <- predict0(fit, newdata, t, bt)
return(result)
}

plot.predict <- function(list0, cols=c("red", "orange", "lightgreen", "lightblue")){
x <- list0[[1]][, 1]
y1 <- rep(1, length(x))
y2 <- 1 - list0[[1]][, 5]
plot(x, y2, xlim = c(0, max(x)), ylim = c(0, 1), type = "s", xlab = "Time after surgery", ylab = "Probability")
line(x, y3, type = "s")
line(x, y4, type = "s", col = "black")
polygon(c(x, rev(x)), c(y3, rev(y2)), col = cols[2])
y5 <- rep(0, length(x))
polygon(c(x, rev(x)), c(y4, rev(y5)), col = cols[4])
}
prob.predict <- function(time, probs){
temp <- tail(probs[probs[, 1] <= 6, , 1]
temp <- c(temp[, 2:3], temp[, 4] + temp[, 5])
names(temp) <- c("P1", "P2", "P3")

49
haz<-function(fit,newdata){
  beta<-summary(fit)$coef[,1]
  beta<-beta[order(names(beta))]
  # # # # # # # # # # # #coefficient# # # # # # # # # # # # # # # # # #
  newdata$trans<-NULL
  ZB<-as.matrix(newdata)%*%beta
  # # # # # # # # # # # #baseline# # # # # # # # # # # # # # # # # # # #
  fit.base<-survfit(fit)
  strata<-fit.base$strata
  inx<-c(cumsum(strata)-strata,sum(strata))
  H.base<-sapply(1:3,function(x)
    cbind(fit.base$time[(inx[x]+1):inx[x+1]],
    -log(fit.base$surv[(inx[x]+1):inx[x+1]])))
  # # # # # # # # # # # #hazard# # # # # # # # # # # # # # # # # # # # # # # #
  H.Z<-sapply(1:3,function(x) cbind(H.base[[x]][,1],H.
    baseline[[x]][,2]*exp(ZB[x]),x))
  hazard<-rbind(H.Z[[1]],H.Z[[2]],H.Z[[3]])
  hazard<-as.data.frame(hazard)
  colnames(hazard)<-c("time","Haz","trans")
  return(hazard)
}
B Codes for the Analysis

```r
# new data

covs <- c("periop", "surgery", "tusi", "nodal", "age50", "adjchem")
(lvls <- getlevel(covs, bc.long))
y <- data.frame(periop = "no periop chemo", surgery = "mastectomy with RT", tuvi = ">5 cm",
    nodal = "node positive", adjchem = "no adj chemo",
    age50 = "<50")
fit <- coxph(Surv(time, status) ~ trans + strata(trans), bc.long, -c(1:3, 6, 9:15))
newdata <- newprep(y, covs, fit, lvls)

# predict

results <- predict(fit, newdata, t = 0, bt = "Tstart")
plot.predict(results)
legend("bottomleft", c("Surgery to Death", "Surgery to Rec to Death", "Recurrence", "DFS"),
    lwd = 2, col = c("red", "orange", "lightgreen", "lightblue"), cex = 0.6)

# se

tmat <- trans.illdeath()
M <- 100
tvec = results$probs[, 1]
history <- list(state = 1, time = 0, tstates = c(0, 0, 0))
theta <- function(data) {
    dat <- data[, -c(1:3, 6, 9:15)]
    fit <- coxph(Surv(time, status) ~ trans + strata(trans), dat)
    Haz <- haz(fit, newdata)
    bt <- summary(fit)$coef[1]["Tstart"]
    beta.state <- matrix(0, 3, 3)
    beta.state[1, 2] <- bt
    res <- mssample(Haz, trans = tmat, clock = "reset", history = history,
        beta.state = beta.state, tvec = tvec, M = M)
    c(as.matrix(res[, -1]))
}
theta(bc.long)
theta2 <- function(data) {
    fit0 <- coxph(Surv(time, status) ~ trans + strata(trans), data, -c(1:3, 6, 9:15))
}
```

51
temp0 <- predict(fit0,newdata,t=5, bt="Tstart")
probs <- temp0$probs
temp <- c(probs[,2:5])
return(temp)
}
msboot(theta2,bc.long,B=1,id="id")

theta2 <- function(data){
fit0 <- coxph(Surv(time,status)^.-trans+strata(trans), data[-c(1:3,6:9:15)])
res0 <- predict(fit0,newdata,t=12, bt="Tstart")
probs <- res0$probs
temp <- c(as.matrix(probs))
return(temp)
}

plot.predict(results)
lines(res[,1],res[,2]+res[,3],col=2)
lines(res[,1],res[,2],col=2)
theta2 <- function(data, trans=tmat){
fit <- coxph(Surv(time,status)^.-trans+strata(trans), data[-c(1:3,6:9:15)])
results <- predict(fit,newdata,t=0, bt="Tstart")
c(as.matrix(results$probs[,2:5]))
}
res <- msboot(theta,bc.long,B=5,id="id")

M <- 10
tvec=6
history <- list(state=1,time=0,tstates=c(0,0,0))
theta <- function(data){
dat <- data[-c(1:3,6:9:15)]
fit <- coxph(Surv(time,status)^.-trans+strata(trans), dat)
Haz <- haz(fit,newdata)
bet <- summary(fit)$coef[,1]["Tstart"]
beta.state <- matrix(0,3,3)
beta.state[1,2] <- bet
res <- mssample(Haz,trans=tmat,clock="reset",
history=history,beta.state=beta.state,tvec=tvec,M=M)
c(as.matrix(res[,1]))
```r
res <- msboot(theta, bc = long, B = 50, id = "id")
LCI <- apply(res, 1, function(x) quantile(x, probs = 0.025))
UCI <- apply(res, 1, function(x) quantile(x, probs = 0.975))
CI <- cbind(LCI, UCI)

# Extended Markov renewal vs. Markov renewal#

fit2 <- coxph(Surv(time, status) ~ .trans + strata(trans) - Tstart, bc = long[,-c(1:3, 6, 9:15)])
newdata2 <- newdata
newdata2$Tstart <- NULL
results2.b <- predict(fit2, newdata2, t = 0, bt = "Tstart")
results2.a <- predict(fit2, newdata2, t = 2, bt = NULL)
results.b <- predict(fit, newdata, t = 0, bt = "Tstart")
results.a <- predict(fit, newdata, t = 2, bt = "Tstart")
par(mfrow = c(1, 1))
plot.predict(results2.a)
legend("bottomleft", c("Surgery to Death", "Surgery to Rec to Death", "Recurrence", "DFS"),
  lwd = 2, col = c("red", "orange", "lightgreen", "lightblue"), cex = 0.75)
plot.predict(results.b)
legend("bottomleft", c("Surgery to Death", "Surgery to Rec to Death", "Recurrence", "DFS"),
  lwd = 2, col = c("red", "orange", "lightgreen", "lightblue"), cex = 0.75)
plot.predict(results2.a)
legend("bottomleft", c("Surgery to Death", "Surgery to Rec to Death", "Recurrence", "DFS"),
  lwd = 2, col = c("red", "orange", "lightgreen", "lightblue"), cex = 0.75)
plot.predict(results.a)
legend("bottomleft", c("Surgery to Death", "Surgery to Rec to Death", "Recurrence", "DFS"),
  lwd = 2, col = c("red", "orange", "lightgreen", "lightblue"), cex = 0.75)

# For patient c

results.c <- predict1(fit, newdata, t = 5, bt = "Tstart", r = 0.5)
```
results.c2 <- predict1(fit, newdata, t=5, bt="Tstart", r=1)
results.c3 <- predict1(fit, newdata, t=5, bt="Tstart", r=2)
results.c4 <- predict1(fit, newdata, t=5, bt="Tstart", r=4)

plot(results.c[,1:2], ylim=c(0,1), type="l", xlab="Time after surgery", ylab="Survival Probability")
lines(results.c2[,1:2], type="l", col=2,lwd=2)
lines(results.c3[,1:2], type="l", col=3,lwd=2)
lines(results.c4[,1:2], type="l", col=4,lwd=2)
legend("topright", c("Reccurence after 0.5 year","Reccurence after 1 year","Reccurence after 2 years","Reccurence after 4 years"), lwd=2, col=1:4)

results2.c <- predict1(fit2, newdata2, t=5, bt=NULL, r=0.5)
results2.c2 <- predict1(fit2, newdata2, t=5, bt=NULL, r=1)
results2.c3 <- predict1(fit, newdata, t=5, bt="Tstart", r=2)
plot(results2.c[,1:2], ylim=c(0,1), type="l", xlab="Time after surgery", ylab="Survival Probability")
lines(results2.c2[,1:2], type="l", col=2)

#### CIs by simulation

res2 <- msboot(theta2, bc.long, B=500, id="id")
save(res, file="boot.alltime.M100.B500")
load("boot.alltime.M100.B500")
se <- apply(res, 1, sd)
LCI <- apply(res, 1, function(x) quantile(x, probs=0.025))
LCI <- matrix(LCI, ncol=3)
UCI <- apply(res, 1, function(x) quantile(x, probs=0.975))
UCI <- matrix(UCI, ncol=3)
head(UCI)
tvec <- results.b[[1]][,1]
plot(results.b[[1]][,1], results.b[[1]][,2], type="l", ylim=c(0,1)
      , xlab="Time after surgery", ylab="Probability", main="DFS")
lines(tvec, LCI[,1], type="l", lty=3)
lines(tvec, UCI[,1], type="l", lty=3)
plot(results.b[[1]][,1], results.b[[1]][,3], type="l", ylim=c(0,1),
    xlab="Time after surgery", ylab="Probability", main="Reccurrence")
lines(tvec, LCI[,2], type="l", lty=3)
lines(tvec, UCI[,2], type="l", lty=3)

plot(results.b[[1]][,1], (results.b[[1]][,4] + results.b[[1]][,5]),
    type="l", ylim=c(0,1),
    xlab="Time after surgery", ylab="Probability", main="Death")
lines(tvec, LCI[,3], type="l", lty=3)
lines(tvec, UCI[,3], type="l", lty=3)

fit <- coxph(Surv(time, status) ~ . - trans + strata(trans), bc.long[, -c(1:3, 6:9:15)])

# For patient c
results.c <- predict1(fit, newdata, t=5, bt="Tstart", r=0.5)
results.c2 <- predict1(fit, newdata, t=5, bt="Tstart", r=1)
results.c3 <- predict1(fit, newdata, t=5, bt="Tstart", r=2)
results.c4 <- predict1(fit, newdata, t=5, bt="Tstart", r=4)
plot(results.c[,1:2], ylim=c(0,1), type="l", xlim=c(5, 12.4), col=1)
lines(results.c2[,1:2], type="l", col=2)
lines(results.c3[,1:2], type="l", col=3)
lines(results.c4[,1:2], type="l", col=4)

results2.c <- predict1(fit2, newdata2, t=5, bt=NULL, r=0.5)
results2.c2 <- predict1(fit2, newdata2, t=5, bt=NULL, r=1)
results2.c3 <- predict1(fit, newdata, t=5, bt="Tstart", r=2)
plot(results2.c[,1:2], ylim=c(0,1), type="l", xlim=c(5, 12.4), col=1)
lines(results2.c2[,1:2], type="l", col=2)

setwd("/Users/songoku/Documents/Master Thesis/Syntax")
load("results")
pr <- results$probs
plot(pr[,1], pr[,2], lwd=2, type="l", xlab="Probability", ylab="time after surgery")
lines(pr[,1], pr[,2]+pr[,3], lwd=2, type="l")
lines(pr[,1], pr[,2]+pr[,3]+pr[,4], lwd=2, type="l")
pr2<-results2$a$probs
lines(pr[,1], pr[,2], lwd=2, type="l", col=2)
lines(pr[,1], pr[,2]+pr[,3], lwd=2, type="l", col=2)
lines(pr[,1], pr[,2]+pr[,3]+pr[,4], lwd=2, type="l", col=2)
legend("bottomleft", lwd=2, c("Extended Markov renewal","Markov renewal"), col=1:2, cex=0.75)