A general piecewise multi-state survival model: Application to breast cancer

- Juan Eloy Ruiz-Castro, Mariangela Zenga
- A general piecewise multi-state survival model: Application to breast cancer
- *Statistical Methods & Applications, vol. 29, pp. 813-843*
- DOI: https://doi.org/10.1007/s10260-019-00505-6

A general piecewise multi-state survival model: Application to breast cancer

Juan Eloy Ruiz-Castro[1] and Mariangela Zenga[2]

[1]Department of Statistics and Operational Research and IEMath-GR. University of Granada. Faculty of Science. Campus Fuentenueva s/n. 18071, Spain. e-mail: jeloy@ugr.es. Phone: +34 958 243 712.

[2]Department of Statistics and Quantitative Methods, University of Milano-Bicocca. Via Bicocca degli Arcimboldi, 8, 20126, Milano, Italy. e-mail: mariangela.zenga@unimib.it. Phone: +39 026 448 3158.

ABSTRACT

Multi-state models are considered in the field of survival analysis for modelling illnesses that evolve through several stages over time. Multi-state models can be developed by applying several techniques, such as non-parametric, semi-parametric and stochastic processes, particularly Markov processes. When the development of an illness is being analysed, its progression is tracked periodically. Medical reviews take place at discrete times, and a panel data analysis can be formed. In this paper, a discrete-time piecewise non-homogeneous Markov process is constructed for modelling and analysing a multi-state illness with a general number of states. The model is built, and relevant measures, such as survival function, transition probabilities, mean total times spent in a group of states and the conditional probability of state change, are determined. A likelihood function is built to estimate the parameters and the general number of cut-points included in the model. Time-dependent covariates are introduced, the results are obtained in a matrix algebraic form and the algorithms are shown. The model is applied to analyse the behaviour of breast cancer. A study of the relapse and survival times of 300 breast cancer patients who have undergone mastectomy is developed. The results of this paper are implemented computationally with MATLAB and R.

Keywords: Survival, breast cancer, piecewise Markov model, multi-state model

1. Introduction

In survival analysis, it is usual to model the progression of a disease that can occupy several states or ‘stages’ over time (e.g., alive without disease, local recurrence, distant metastasis, death). Studies analysing complex disease behaviour often use multi-state models that have been extensively developed over the last few decades (Andersen and Keiding, 2001; Hougaard, 1999; Putter et al., 2007; Meira-Machado et al., 2009). Several methodologies have been used to analyse parametric and non-parametric multi-state models. In a multi-state framework, Cortese and Andersen (2010) presented some approaches for estimating cumulative incidences and survival probabilities when internal time-dependent covariates are included. Chen et al. (2010) were concerned with the analysis of data from progressive multi-state disease processes in which individuals are scheduled to be seen at periodic, pre-scheduled assessment times. Farewell and Tom (2014) showed that the multi-state modelling approach provides a convenient framework for handling a wide variety of medical applications characterized by multiple events and longitudinal data. Recently, flexible multi-state models for serial hospital admissions and death in heart failure of patients who are able to accommodate important features of disease progression, such as multiple ordered events and the competing risks of death and hospitalization, were proposed by Ieva et al. (2015).

Several methodologies have been considered to study multi-state models. The usefulness of Markov modelling techniques to analyse multi-state models has previously been demonstrated. Several papers have described disease progression over time using a multi-state model that introduces assumptions of homogeneity. This hypothesis assumes that the transition rates between any two different states of an illness are constant over time. Pérez-Ocón et al. (1998) analysed the evolution of breast cancer using a continuous-time homogeneous Markov model. Jackson (2011) reviewed Markov models and their extensions, which can be fitted to panel-observed data. The homogeneity hypothesis sometimes appears unrealistic, however, given that the evolution of the disease is not constant over time. To overcome this problem, Pérez-Ocón et al. (2001) proposed a non-homogeneous approach using a piecewise Markov process, with the transition intensity functions being step functions. The authors also analysed the survival evolution of different groups of patients by introducing covariates into the model. Commenges and Joly (2004) considered a multi-state model, with five states, for jointly modelling dementia, institutionalization and death to analyse the relationships between these factors through a non-homogeneous Markov model.

Continuous-time Markov models have found wide application in medicine and social sciences for studying data that record life history events for individuals (Kalbfleisch and Lawless, 1985; Jackson et al., 2003). Santamaría et al. (2009) considered a homogeneous Markov model to analyse the behaviour of bladder carcinoma. The importance of modelling in continuous time is evident in survival analysis, but discrete time also plays an important role in biomedical analysis. Many studies apply several methodologies to show the importance of discrete time (see, for example, Van Den Hout 2016). In general, models in discrete time are treated similarly to models in continuous time, with piecewise-constant main probabilistic functions. Moreover, Singer and Willett (2003) developed multilevel models for individual change and hazard/survival models for event occurrence (in both discrete and continuous time). Bacchetti et al. (2010) considered a discrete-time, multi-state model to analyse the progression of liver fibrosis due to hepatitis C following liver transplantation. When the development of an illness is being analysed, its progression is tracked periodically form. In this way, medical revisions take place at discrete times, and a panel data analysis can be built. It can be interesting to know the state of an illness at a certain revision. In addition to continuous time, discrete time could then be introduced to analyse the progression of a disease, particularly with discrete-time, multi-state models when the evolution passes through different states. The structure of the models and the associated measures should be studied with different techniques.

When the embedded lifetime distributions in the evolution of a disease are complex, such as transition times between any two states, intractable expressions sometimes appear in the modelling and in the associated measures. It is interesting to work out the modelling and results algorithmically and computationally, which enables us to obtain expressions that can be computationally applied and more easily implemented. One class of distributions that has a matrix structure and provides an algorithmic Markovian methodology is the phase-type distribution (Neuts, 1981). Titman (2014) developed an alternative application of phase-type distributions for semi-Markov models to provide a computationally tractable approximate likelihood. The matrix structure provided by Markov processes makes it possible to express the modelling in this way. A well-structured matrix form enables us to express different models in a similar way depending on the inner structure of the matrices. The matrix expressions are the same, but different transitions inside the matrices can lead to different models.

This paper provides several novelties from a theoretical and medical point of view. In terms of theory, 1) the evolution time of a general disease with a general number of states when it is observed in discrete time is analysed by a multi-state model, and a non-homogeneity is introduced in a discrete multi-state model through a discrete-time piecewise non-homogeneous Markov process; 2) following the good properties of the phase-type distribution class, the model is developed in matrix and algorithmic forms to be applied to any general case; 3) a first approximation to time-dependent covariate vectors is introduced; 4) the likelihood function for estimating the parameters and cut-points of the model is built. Both have been estimated jointly by taking into account that the cut-points are non-negative integer values; 5) the lifetime distribution and several important associated measures, such as the mean total times, first-passage time distributions and conditional probabilities of state change, are determined in a well-structured form.

From a medical point of view, the model developed is applied to a cohort of 300 patients to analyse the progress of breast cancer under different treatments with interesting conclusions. All patients have similar features (all patients had some axillary nodes affected, and certain basic criteria were adopted in the application of treatments) and can move through three different states by time: State 1 (patient operated with no signs of disease), the initial state for all patients after mastectomy; State 2 (relapse), the state at which the tumour has a local regional recurrence in the same sampling location as the initial tumour, at the site of the scar from the initial operation, in the supraclavicular or axillary ganglionic regions, or in the internal mammary chain; or State 3, the state of death, reached when a patient has died as a consequence of the initial breast cancer (with or without local relapse). Initially, all patients are in State 1; the malignant tumour has been removed. From a medical point of view, treatment effectiveness is analysed considering the measures developed in this work. New prognosis factors, such as the number of infected glands and menopausal state, are included in the paper jointly in the treatments. A goodness-of-fit analysis is carried out. It is important to highlight that the developed methodology is not an immediate consequence of the continuous case.

The remainder of this paper is organized as follows. In Section 2, the piecewise Markov model in discrete time is presented. Section 3 describes the likelihood function to estimate the parameters of the new model, while Section 4 gives some interesting measures associated with the described model. In Section 5, the breast cancer data used is described, while in Section 6 the methodology is applied to analyse the behaviour of breast cancer. Section 7 contains some concluding remarks. The modelling and the results are implemented computationally using MATLAB and R.

2. The Markov model

The usefulness of Markov models has been indicated in disease progression modelling. It is common in the survival literature to consider that the behaviour of the process is homogeneous over time, i.e., the transition probability is the same after each step on the real line. Sometimes this assumption is unrealistic, given that a disease behaves differently at different times. In these cases, a non-homogeneous model should be considered. One approximation of a non-homogeneous model may be obtained in the following way.

2.1. The piecewise model

Let $\{X_\nu\,; \nu \geq 0\}$ be a discrete-time Markov process with a state space $S = \{1, 2, ... r, r + 1\}$, where $\{1, 2, ..., r\}$ are transient states of the process and the state $r+1$ is an absorbing state. Moreover, the absorbing state can be changed by an absorbing class. The initial distribution for the transient states is denoted by $\boldsymbol{\alpha}$.

We assume that the progression of a disease varies over time and is observed at discrete times. The non-negative real line is partitioned through $k-1$ positive integer cut-points,

$$c_0=0 <\mathrm{c}_1<\ldots<c_{k-1}<c_k=\infty,$$

which define k intervals of time $I_l = \left[c_{l-1}, c_l\right);\ l = 1,\ldots,k$. The length of each interval is given by

$$a_l = c_l - c_{l-1}; \qquad l = 1,\ldots,k-1.$$

We assume that the behaviour of the disease in each time interval can be modelled by a homogeneous Markov process. If the *l-th* interval is considered, the transitions are governed by a local transition probability matrix given by

$$\mathbf{P}_l = \left(\begin{array}{c|c} \mathbf{T}_l & \mathbf{T}_l^0 \\ \hline 0 & 1 \end{array}\right), \quad l = 1,\ldots k,$$

where the matrix $\mathbf{T}_l$ contains the transition probabilities between any two transient states in the interval I_l and $\mathbf{T}_l^0$ is a column vector whose elements are the probabilities of reaching the absorbing state from any transient state inside the interval I_l.

2.2. Transition probabilities for the piecewise model

The transition probability matrix for the proposed piecewise model is built in matrix and algorithmic forms from the local transition probability matrices. Given that at time n, the disease is in state i, then the probability of being in state j at time m is given by the element (i, j) of the following matrix:

$$\mathbf{P}(n,m) = \left(\begin{array}{c|c} \mathbf{T}(n,m) & \mathbf{T}^0(n,m) \\ \hline 0 & 1 \end{array}\right) \quad . \tag{1}$$

This matrix is also expressed by blocks according to the class of the states. The matrix $\mathbf{T}(n, m)$ contains the transition probabilities between any two transient states, and $\mathbf{T}^0(n, m)$ is a column vector that contains the transition probabilities from a transient state at time n up to an absorbing state at time *m*. This matrix is calculated by applying the Chapman-Kolmogorov equation at any cut-point between *n* and *m* and by considering the matrix block structure. Thus,

$$\mathbf{T}(n,m)=\begin{cases}\mathbf{T}_{l_1}^{m-n} & ;\quad m,n\in I_{l_1}\\ \mathbf{T}_{l_1}^{c_{l_1}-n}\mathbf{T}_{l_1+1}^{m-c_{l_1}} & ;\quad n\in I_{l_1}, m\in I_{l_1+1}\\ \mathbf{T}_{l_1}^{c_{l_1}-n}\prod_{i=1}^{d-1}\mathbf{T}_{l_1+i}^{a_{l_1+i}}\mathbf{T}_{l_2}^{m-c_{l_2-1}} & ;\quad n\in I_{l_1}, m\in I_{l_2}, d=l_2-l_1>1.\end{cases} \tag{2}$$

Given that matrix (1) is stochastic, then $\mathbf{T}^0(n,m)=\mathbf{e}-\mathbf{T}(n,m)\mathbf{e}$, being **e** a column vector of ones with appropriate order throughout the paper.

Therefore, the probability that at time *m* the disease occupies state *j*, given that it is in state *i* at time *n*, is given by the element (*i, j*) of the matrix **P**(*n, m*),

$$p_{i,j}(n,m)=P(X_m=j\mid X_n=i)=(\mathbf{P}(n,m))_{ij}\ .$$

The transient distribution vector is given by

$$\mathbf{p}(m)=(P(X_m=i))_{i\in S}=(\boldsymbol{\alpha},0)\mathbf{P}(0,m),$$

where $\boldsymbol{\alpha}$ is the initial distribution for the transient states.

2.3. Covariates

We assume that in each interval of time I_l, a covariate column vector $\mathbf{z}_l$ is introduced. A first approximation of the time-dependent covariates is introduced as follows. The covariate vector is introduced in the piecewise model for the transition between states *i* and *j* in the interval I_l as

$$[\mathbf{T}_l(\mathbf{z}_l)]_{ij}=[\mathbf{T}_l]_{ij}\exp\{\mathbf{z}_l^{'}\boldsymbol{\beta}_{ij}^{l}\}\ ,\ l=1,\ldots,k;\ i,j=1,\ \ldots,\ r,$$

where $\boldsymbol{\beta}_{ij}^{l}$ is the column vector of regression coefficients for the transition between states *i* and *j* in the interval I_l and $[\mathbf{T}_l]_{ij}$ is the baseline transition probability between states *i* and *j* in the interval of time I_l. Therefore, it follows that $[\mathbf{T}_l(\mathbf{z}_l)]_{ij}$ is the transition probability for one step in the interval I_l for the subjects characterized by the vector $\mathbf{z}_l$. The scalar product of vectors $\mathbf{z}_l^{'}\boldsymbol{\beta}_{ij}^{l}$ can be interpreted as in the Cox model, where $\boldsymbol{\beta}_{ij}^{l}$ represents the regression coefficients of the covariates on transitional probabilities. Thus, in this interval of time, the probability ratio for items *a* and *b* with covariate vectors $\mathbf{z}_l^a$ and $\mathbf{z}_l^b$, respectively, for the transition *i*→*j* is given by

$$\frac{\left[\mathbf{T}_l\left(\mathbf{z}_l^a\right)\right]_{ij}}{\left[\mathbf{T}_l\left(\mathbf{z}_l^b\right)\right]_{ij}} = \exp\left\{\left(\mathbf{z}_l^a - \mathbf{z}_l^b\right)' \boldsymbol{\beta}_{ij}^l\right\}.$$

Given the covariate vector $\mathbf{z}_l$ in the interval of time I_l, the transition probability matrix is given by

$$\mathbf{P}_l\left(\mathbf{z}_l\right) = \left(\begin{array}{c|c} \mathbf{T}_l\left(\mathbf{z}_l\right) & \mathbf{T}_l^0\left(\mathbf{z}_l\right) \\ \hline 0 & 1 \end{array}\right), \quad l = 1, \ldots, k,$$

and from these, the piecewise transition probabilities, depending on the covariate vectors, can be worked out similarly to (2). Consequently,

$$\mathbf{P}\left(n, m; \mathbf{z}_{l_1}, \ldots, \mathbf{z}_{l_2}\right) = \left(\begin{array}{c|c} \mathbf{T}\left(n, m; \mathbf{z}_{l_1}, \ldots, \mathbf{z}_{l_2}\right) & \mathbf{T}^0\left(n, m; \mathbf{z}_{l_1}, \ldots, \mathbf{z}_{l_2}\right) \\ \hline \mathbf{0} & 1 \end{array}\right), \tag{3}$$

and the transition probability between states *i* at time *n* and states *j* at time *m* is denoted as

$$p_{i,j}\left(n, m; \mathbf{z}_{l_1}, \ldots, \mathbf{z}_{l_2}\right) = \left(\mathbf{P}\left(n, m; \mathbf{z}_{l_1}, \ldots, \mathbf{z}_{l_2}\right)\right)_{ij},$$

where $n \in I_{l_1}$ and $m \in I_{l_2}$.

3. Likelihood function

The parameters of the model are estimated by maximum likelihood. We assume that *n* items are observed, all beginning in state 1, and item *i* is observed at m_i change times, the last time being death or censorship. Given that the item is observed at different change times, then for any item, the value of the covariate vector and the corresponding state is observed. Therefore, a sequence of times, states and values of the covariate vector is achieved for each item *i*: $0 = t_{i,1} < t_{i,2} < \cdots < t_{i,m_i}$, $1 = x_1^i, \ldots, x_{m_i}^i$ and $\mathbf{z}_{l_1}^i, \ldots, \mathbf{z}_{l_{m_i}}^i$, respectively. $\mathbf{z}_{l_s}^i$ corresponds to the covariate vector for the interval that contains the time $t_{i,s}$ for item *i* and for $s = 1, \ldots, m_i$.

As such, if a patient is observed at times $t_{i,1} \in I_l$ and $t_{i,2} \in I_{l+a}$, it contributes to the likelihood function with the factor

$$f_{x_1^i}^{l}\left(c_l - t_{i,1}, \mathbf{z}_l^i; \mathbf{T}_l, \boldsymbol{\beta}^l\right) \prod_{u=l+1}^{l+a-1} f_{x_1^i}^{u}\left(c_u - c_{u-1}, \mathbf{z}_u^i; \mathbf{T}_u, \boldsymbol{\beta}^u\right) f_{x_1^i}^{l+a}\left(t_{i,2} - c_u - 1, \mathbf{z}_{l+a}^i; \mathbf{T}_{l+a}, \boldsymbol{\beta}^{l+a}\right) T_{x_1^i, x_2^i}^{l+a}\left(\mathbf{z}_{l+a}^i\right),$$

where $f_x^q\left(t, \mathbf{z}; \mathbf{T}_q, \boldsymbol{\beta}^q\right)$ is the sojourn time probability in state x at time t in the interval I_l, defined as

$$f_x^q\left(t, \mathbf{z}; \mathbf{T}_q, \boldsymbol{\beta}^q\right) = \left[T_{x,x}^q\left(\mathbf{z}\right)\right]^t,$$

where $T_{x,y}^q\left(\mathbf{z}\right)$ is the element (x, y) of the matrix $\mathbf{T}_q(\mathbf{z})$.

The parameters of the model and the cut-points are estimated through maximum likelihood by considering that the matrices $\mathbf{P}_q\left(\mathbf{z}_q^i\right)$ are stochastic matrices for any value of the covariate vector and any item i.

The likelihood function has been built to estimate the parameters and cut-points. It is described in Appendix A.

4. Measures

Some interesting measures associated with the model described above have been obtained. The survival function, mean time to absorption, mean total time in a given state up to a certain time, first-passage time distribution and some conditional probabilities have been worked out in a matrix algebraic form. The measures are introduced without covariates to facilitate the expressions; the case with covariates is analogous.

4.1. Lifetime distribution and mean time up to absorption

The distribution of time to absorption can be expressed in matrix form by considering the survival function. The discrete survival function is the probability that at time m, the disease occupies a transient state. It is given by

$$S\left(m\right) = \boldsymbol{\alpha}\mathbf{T}\left(0, m\right)\mathbf{e}, \quad m = 0, 1, 2, \ldots$$

Given the survival function, the mean time to absorption can be calculated from the expressions given in (2) as

$$\mu = \sum_{m=0}^{\infty} S(m) = \boldsymbol{\alpha} \sum_{\mathbf{m}=\mathbf{0}}^{\infty} \mathbf{T}(0,m)\mathbf{e}$$

$$= \boldsymbol{\alpha}\left[(\mathbf{I}-\mathbf{T}_1)^{-1}\left(\mathbf{I}-\mathbf{T}_1^{c_1}\right) + \sum_{j=1}^{k-2}\prod_{i=1}^{j}\mathbf{T}_i^{a_i}\left(\mathbf{I}-\mathbf{T}_{j+1}\right)^{-1}\left(\mathbf{I}-\mathbf{T}_{j+1}^{a_{j+1}}\right) + \prod_{i=1}^{k-1}\mathbf{T}_i^{a_i}\left(\mathbf{I}-\mathbf{T}_k\right)^{-1}\right]\mathbf{e}.$$

4.2. Mean total time in each transient state

In a multi-state model, a disease can pass through different states over time. The mean total time in each transient state up to a certain time $\nu \in I_l$ can be determined in matrix form as

$$G(\nu) = \boldsymbol{\alpha}\sum_{m=0}^{\nu}\mathbf{T}(0,m)$$

$$= \boldsymbol{\alpha}\left[(\mathbf{I}-\mathbf{T}_1)^{-1}\left(\mathbf{I}-\mathbf{T}_1^{c_1}\right) + \sum_{j=1}^{l-2}\prod_{i=1}^{j}\mathbf{T}_i^{a_i}\left(\mathbf{I}-\mathbf{T}_{j+1}\right)^{-1}\left(\mathbf{I}-\mathbf{T}_{j+1}^{a_{j+1}}\right) + \prod_{i=1}^{l-1}\mathbf{T}_i^{a_i}\left(\mathbf{I}-\mathbf{T}_l\right)^{-1}\left(\mathbf{I}-\mathbf{T}_l^{\nu-c_{l-1}+1}\right)\right].$$

If ν tends to infinity, then the mean total time in each transient state before absorption is equal to

$$\mathbf{G} = \boldsymbol{\alpha}\left[(\mathbf{I}-\mathbf{T}_1)^{-1}\left(\mathbf{I}-\mathbf{T}_1^{c_1}\right) + \sum_{j=1}^{k-2}\prod_{i=1}^{j}\mathbf{T}_i^{a_i}\left(\mathbf{I}-\mathbf{T}_{j+1}\right)^{-1}\left(\mathbf{I}-\mathbf{T}_{j+1}^{a_{j+1}}\right) + \prod_{i=1}^{k-1}\mathbf{T}_i^{a_i}\left(\mathbf{I}-\mathbf{T}_k\right)^{-1}\right].$$

4.3. First-passage time distribution

Another interesting distribution that should be estimated is the first-passage time, defined as the probability that the disease occupies the transient state *j* for the first time at a given time *m*. In the progression of a disease, which can pass through several transient states before absorption, it is important to calculate the probability that the disease progression enters at a determinate level at a certain time.

The probability mass function of the first-passage time for the transient state *j* at time *m* is given by

$$f_j(m) = \boldsymbol{\alpha}_{-j}\mathbf{T}_{-j}^{*}(0,m-1)\mathbf{T}_l^{(j)}, \text{ for } j = 1,\ldots, r, \text{ and } m \in I_l,$$

where the vector $\boldsymbol{\alpha}_{-j}$ is the initial distribution without the *j-th* element, $\mathbf{T}_l^{(j)}$ is the *j-th* column of the matrix $\mathbf{T}_l$ without the *j-th* row and $\mathbf{T}_{-j}^{*}(\cdot,\cdot)$ is given by

$$\mathbf{T}^{*}_{-j}\left(0,k\right)=\begin{cases}\mathbf{T}^{k}_{1,-j} & ; \quad k\in I_1\\ \mathbf{T}^{c_1}_{1,-j}\mathbf{T}^{k-c_1}_{2,-j} & ; \quad k\in I_2\\ \mathbf{T}^{c_1}_{1,-j}\prod_{i=1}^{s-2}\mathbf{T}^{a_{i+1}}_{i+1,-j}\mathbf{T}^{k-c_{s-1}}_{s,-j} & ; \quad k\in I_s, s>2,\end{cases}$$

where the matrix $\mathbf{T}_{l,-j}$ is the matrix $\mathbf{T}_l$ without the *j-th* row and column. This is a defective distribution given that the probability of reaching transient state *j* is lower than one (in other words, the absorption can have occurred previously).

If *m* tends to infinity, then this measure is the probability of reaching state *j*, and it is given by

$$\sum_{m=1}^{\infty} f_j\left(m\right)=\boldsymbol{\alpha}_{-j}\left[\left(\mathbf{I}-\mathbf{T}_{1,-j}\right)^{-1}\left(\mathbf{I}-\mathbf{T}^{c_1}_{1,-j}\right)+\sum_{s=1}^{k-2}\prod_{i=1}^{s}\mathbf{T}^{a_i}_{i,-j}\left(\mathbf{I}-\mathbf{T}_{s+1,-j}\right)^{-1}\left(\mathbf{I}-\mathbf{T}^{a_{s+1}}_{s+1,-j}\right)\right.$$
$$\left.+\prod_{i=1}^{k-1}\mathbf{T}^{a_i}_{i,-j}\left(\mathbf{I}-\mathbf{T}_{k,-j}\right)^{-1}\right]\mathbf{T}^{(j)}_{l}.$$

5. Application to breast cancer: description of the data set.

The Department of Radiology at the Hospital Clínico of Granada (Spain) supplied a data set on a cohort of patients who underwent surgical treatment for breast cancer. A total of 300 patients were seen every month between December 1973 and December 1995. All patients were diagnosed and underwent mastectomy in the Hospital Clínico of Granada. Each patient entered the study the month after undergoing mastectomy. They were observed periodically at discrete times; the unit of time considered in this application is one month. The last observed time for each patient is either a censoring time or a completion time. If the patient was alive at the end of the study time (December 1995), the patient is considered censored. All 300 patients were selected because they shared similar characteristics and had at least one axillary node affected. After surgery, these patients might have undergone different forms of treatment: radiotherapy (RT), hormonal therapy (HT), chemotherapy (CT) or combinations of the three. Initially, the treatments were preventive, and all patients with similar symptoms had the same type of treatment. Thus, hormonal therapy was given only to patients with positive oestrogen receptors. Patients receiving chemotherapy reatment were administered three different injected drugs: cyclophosphamide, methotrexate, and 5-fluorouracil. Thirty-nine patients (13% of the total) received all three types of treatment, while 22 patients (7.33%) received a combination of RT and HT. The largest group was the 110 patients (36.67%) receiving RT and CT, while the smallest was the group of 7 patients (2.33%) who received only HT. Table 1 shows the complete distribution of the 300 patients by treatment.

The patients could potentially pass through three states: State 1, no relapse (operation with no signs of disease), State 2, relapse (local regional recurrence of tumours), and State 3, death (as a result of the original disease). A censored patient is a patient who died as a result of other causes (not related to the original tumour), with whom contact had been lost, or who was still alive at the end of the observation period (December 1995). The initial state for all patients is state 1. The mean age of the patients was 52.48, with a standard deviation of 11.02. The minimum and maximum values observed were 25 and 80, respectively; the first quartile was 45, and the third quartile was 60. The total number of censored patients was 122 (40.67%), 110 from no relapse and 12 from relapse.

Previously, with Markovian analysis, the behaviour of breast cancer was analysed by using several methodologies: parametric, non-parametric and semi-parametric (Cox model). The behaviour of the empirical and cumulative hazard functions from State 1 to relapse and death and from relapse to death were studied. Figure 1 shows the empirical cumulative hazard rate to death. We can observe that the behaviour of breast cancer is different at the beginning after surgical treatment, during the middle period and at the end of the follow-up period. This fact was analysed with the medical team, and it was concluded, without formal estimation, that the behaviour of the illness could be divided into three different periods~~.~~ The cut-points of this division will be formally estimated by the maximum likelihood method in Section 6.2.

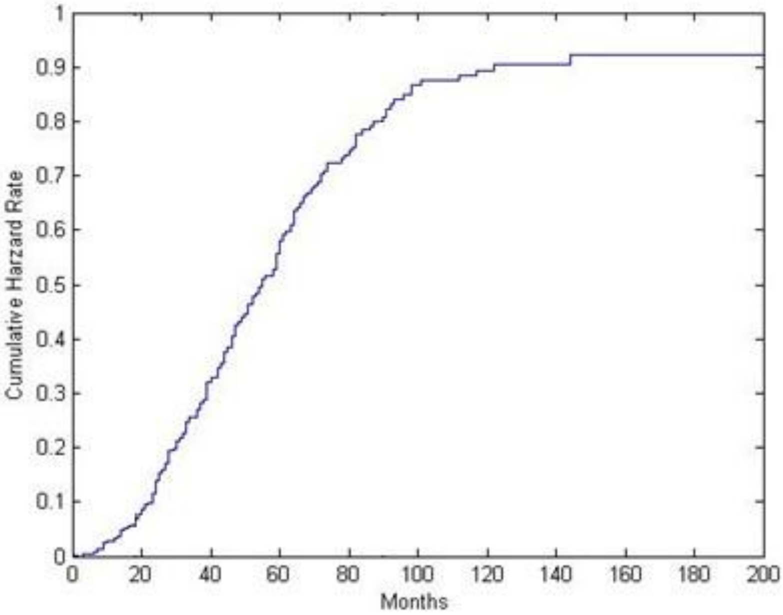

Figure 1. Cumulative hazard rate to death

6. The Markov model

Here, the methodology developed in Section 4 for any number of states and cut-points has been applied to analyse the behaviour of breast cancer. The cut-points and parameters were estimated with the maximum likelihood method by applying the likelihood function described in the Appendix. The model was built and used to determine some interesting measures.

The progression of breast cancer has been modelled using a piecewise Markov model with two cut-points. The Markov assumption was tested using a Cox model (Kalbfleisch and Prentice, 1980). The *p-value* calculated from the data set was 0.1119; thus, there is no empirical evidence for rejecting the null hypothesis of the Markov assumption. This disease can pass through two transient states: state 1, no relapse, and state 2, relapse. The state space is S = {1, 2, 3}, where state 3 is the absorbing state, death. The death state was reached if a patient died as a result of the initial breast cancer. The possible transitions between any two states are given in Figure 2. This figure also shows the number of patients in each transition and the number of patients censored in each state.

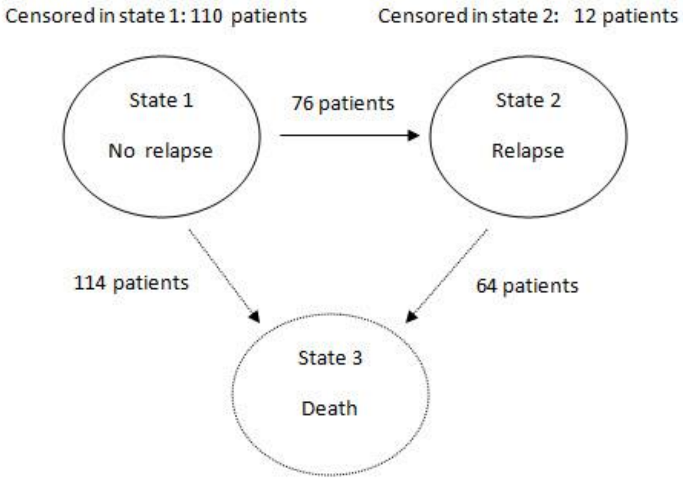

Figure 2. Transitions among states in the piecewise Markov model

In the analysis of the model, the treatments are introduced through a covariate three-vector $\mathbf{z}$'=(RT, HT, CT), with z_h as a dichotomous variable for h = 1, 2, 3 equal to 1 if the corresponding treatment is not applied and 0 if it is. In this case, the vector $\mathbf{z}$ is time-independent, $\mathbf{z}_l = \mathbf{z}$ for l = 1, ..., k. The effect of treatment h on the transition $i \rightarrow j$ is measured in each interval of time by the coefficient vector $\boldsymbol{\beta}_{ij}^{l}$ for l = 1, 2, 3. The transition probability matrix for each interval of time is given by

$$\mathbf{P}_l(\mathbf{z})=\left(\begin{array}{c|c} \mathbf{T}_l(\mathbf{z})=\begin{pmatrix} 1-\dfrac{\lambda_{12}^l}{1+\lambda_{12}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{12}^l}-\dfrac{\lambda_{13}^l}{1+\lambda_{13}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{13}^l} & \dfrac{\lambda_{12}^l}{1+\lambda_{12}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{12}^l} \\ 0 & 1-\dfrac{\lambda_{23}^l}{1+\lambda_{23}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{23}^l} \end{pmatrix} & \mathbf{T}_l^0=\begin{pmatrix} \dfrac{\lambda_{13}^l}{1+\lambda_{13}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{13}^l} \\ \dfrac{\lambda_{23}^l}{1+\lambda_{23}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{23}^l} \end{pmatrix} \\ \hline \mathbf{0} & 1 \end{array}\right),$$

for l = 1, 2, 3, and the transition probabilities, $\mathbf{P}(n,m;\mathbf{z})$, are obtained by considering (3). As shown in Section 2.3, the transition probabilities can be interpreted as in the Cox model. The baseline transition probability can also be interpreted as follows. When the baseline transition probability matrix is considered, i.e., all treatments are applied, $\mathbf{z}$ = (0, 0, 0)', and the transition probabilities $i\rightarrow j$ with $i \neq j$ can be interpreted as

$$\frac{[\mathbf{T}_l]_{ij}}{1-[\mathbf{T}_l]_{ij}}=\lambda_{ij}^l, \qquad \frac{[\mathbf{T}_l^0]_i}{1-[\mathbf{T}_l^0]_i}=\lambda_{i,r+1}^l.$$

The transition probability $i\rightarrow j$ in the interval I_l is λ_{ij}^l times the probability of no occurrence of this transition.

The power in k of the matrix $\mathbf{T}_l(\mathbf{z})$ for l = 1, 2, 3 can be expressed in algorithmic form to optimize the procedure:

$$\mathbf{T}_l^k(\mathbf{z})=\begin{pmatrix} \left[1-\dfrac{\lambda_{12}^l}{1+\lambda_{12}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{12}^l}-\dfrac{\lambda_{13}^l}{1+\lambda_{13}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{13}^l}\right]^k & \displaystyle\sum_{a=1}^{k}\left[1-\dfrac{\lambda_{12}^l}{1+\lambda_{12}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{12}^l}-\dfrac{\lambda_{13}^l}{1+\lambda_{13}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{13}^l}\right]^{k-a}\dfrac{\lambda_{12}^l}{1+\lambda_{12}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{12}^l}\left[1-\dfrac{\lambda_{23}^l}{1+\lambda_{23}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{23}^l}\right]^{a-1} \\ 0 & \left[1-\dfrac{\lambda_{23}^l}{1+\lambda_{23}^l}e^{\mathbf{z}'\boldsymbol{\beta}_{23}^l}\right]^k \end{pmatrix}.$$

The probability that at time m, the disease occupies state j, given that the disease is in state i at time n, when a covariate vector $\mathbf{z}$ is present is given by

$$p_{ij}(n,m;\mathbf{z})=\left(\mathbf{P}(n,m;\mathbf{z})\right)_{ij}.$$

The remaining measures can be determined by considering their corresponding covariates in a similar way.

6.1. Estimates of cut-points

A crucial part of the analysis is represented by the value of the cut-points. As reported in Section 5, the medical team hypothesized that the behaviour of the illness could be divided into three different periods by investigating the cumulative hazard rate to death. This assumption was formally estimated using the Cox phase-type distribution (Faddy, 1998) at the times to death, which estimated three phases. Moreover, the models were estimated for every combination of times, ultimately resulting in $\binom{252}{2} = 31626$ models estimated by using the likelihood function described in Appendix A. The best model (and consequently the value of the optimum cut-points) was detected using the log-likelihood values of the models. The estimated cut-points were 19 and 93 months.

6.2 Estimates of the model with treatments as covariates

Looking at the effects of treatments, it is possible to state that RT treatment was essential in order to avoid relapse. This treatment was applied to 221 patients (73.67%); only 31 (14.03%) underwent a relapse: 9 of them within 18 months, 18 between 18 and 90 months, and 4 after 90 months. These values can be compared with the patients who were not treated with RT. In this case, we had 79 patients with 45 relapses (56.96%): 31 of them before 18 months, 13 between 18 and 90 months, and 1 after 90 months. In this analysis, other treatments were, of course, applied jointly with RT, but when only RT was considered (50 patients, 16.67%), 13 of them had a relapse (26%). Table 1 shows a summary of the data by period and risk group.

To estimate the parameters for the model in equation (3), the likelihood function described in Appendix A was implemented computationally with MATLAB, and the results with their standard errors are shown in Table 2. In this case the parameters estimated are the cut-points, the regression covariate vectors and the parameters inside in matrices $\mathbf{T}_l$, that is, $\lambda_{12}^i, \lambda_{13}^i, \lambda_{23}^i, \boldsymbol{\beta}_{12}^i, \boldsymbol{\beta}_{13}^i, \boldsymbol{\beta}_{23}^i, c_1$ and c_2 for $i = 1, 2, 3$.

The cut-points were estimated to be 19 and 93 months, and the log-likelihood value was −1419.9795. The regression parameter vector can be interpreted through the hazard ratio in a similar way to the Cox model by considering the transition probabilities.

We can observe that the standard error associated with RT treatment in transition 1→3 in the third period of time is very high. For this reason, in this period, RT treatment might not make sense. This fact was contrasted by using the likelihood-ratio test, obtaining *p-value*=0.1911. The null hypothesis was not rejected, and a new model was estimated. The new estimates for the third period are given in Table 3.

		Survivors in state 1		Death from state 1			Censored in state 1		
Treatment	Total	J_1	J_2	J_1	J_2	J_3	J_1	J_2	J_3
RT-HT-CT	39	35	21	2	11	3	0	1	17
RT-HT	22	18	7	3	6	1	0	4	6
RT-CT	110	106	57	2	40	5	0	2	50
HT-CT	12	7	2	1	2	0	0	0	2
RT	50	43	16	3	16	1	0	3	14
HT	7	5	3	0	2	0	0	0	3
CT	13	10	5	0	1	0	0	0	5
No treatment	47	22	4	3	12	0	0	0	3

	Relapse			Survivors in state 2		Death from state 2			Censored in state 2		
Treatment	J_1	J_2	J_3	J_1	J_2	J_1	J_2	J_3	J_1	J_2	J_3
RT-HT-CT	2	2	1	1	1	1	2	0	0	0	2
RT-HT	1	1	0	1	1	0	1	0	0	0	1
RT-CT	2	7	2	2	2	0	7	1	0	0	3
HT-CT	4	3	0	4	1	0	6	1	0	0	0
RT	4	8	1	3	0	1	11	0	0	0	1
HT	2	0	0	2	0	0	2	0	0	0	0
CT	3	4	0	2	0	1	6	0	0	0	0
No treatment	22	6	1	19	5	3	20	1	0	0	5

Table 1. Survivors, death and censored patients for each period of time (J_1≡[0, 18), J_2≡[18, 90), J_3≡ [90, ∞)) and risk group

Parameters	$I_1 \equiv[0, 19)$	$I_2\equiv[19, 93)$	$I_3\equiv[93,\infty)$
λ^l_{12}	0.0017	0.0010	0.0003
(s.e.)	(0.0007)	(0.0005)	(0.0001)
λ^l_{13}	0.0024	0.0062	0.0007
(s.e.)	(0.0013)	(0.0014)	(0.0001)
λ^l_{23}	0.0406	0.0368	0.00005
(s.e.)	(0.0457)	(0.0134)	(0.00000)
$\boldsymbol{\beta}^l_{12}$	(2.3521, –0.0128, 0.6938)	(1.1477, 0.7097, 0.3621)	(0.8288, –0.2252, 1.5149)
(s.e.)	(0.3892, 0.3668, 0.3600)	(0.3963, 0.4878, 0.3813)	(0.9498, 0.7379, 0.8634)
$\boldsymbol{\beta}^l_{13}$	(–0.0396, –0.6451, 1.1945)	(0.2815, 0.1181, 0.2680)	(–22.5351, 0.6529, 0.5781)
(s.e.)	(0.5970, 0.5282, 0.5586)	(0.2765, 0.2363, 0.2209)	(55564.5, 0.3902, 0.6498)
$\boldsymbol{\beta}^l_{23}$	(–0.9735, 0.8986, –0.7378)	(–0.2819, 0.2672, –0.1133)	(5.2433, 4.7289, –5.7713)
(s.e.)	(0.8609, 1.1614, 0.9198)	(0.2885, 0.3583, 0.3054)	(0.4908, 0.4908, 0.6774)

Table 2. Maximum-likelihood estimates and standard errors of baseline transition probabilities and regression coefficients in each interval of time

Parameters	$I_3\equiv[93,\infty)$
λ^l_{12}	0.0003
(s.e.)	(0.00005)
λ^l_{13}	0.0006
(s.e.)	(0.00007)
λ^l_{23}	0.00005
(s.e.)	(0.00000)
$\boldsymbol{\beta}^l_{12}$	(0.8270, –0.2253, 1.5151)
(s.e.)	(0.9499, 0.7379, 0.8635)
$\boldsymbol{\beta}^l_{13}$	(0, 0.6426, 0.3552)
(s.e.)	(—, 0.3937, 0.6316)
$\boldsymbol{\beta}^l_{23}$	(5.2435, 4.7291, –5.7715)
(s.e.)	(0.4908, 0.4908, 0.6774)

Table 3. Maximum-likelihood estimates and standard errors of baseline transition probabilities and regression coefficients in the third interval of time for the reduced model

The conversion factor, if a covariate vector **z** is applied with respect to the risk group with no treatments, is therefore given by $e^{[\mathbf{z}-(1,1,1)]'\boldsymbol{\beta}}$. Table 4 shows the estimated conversion factor with respect to no treatments for each interval of time and each transition according to the type of treatment. We can observe the importance of radiotherapy for avoiding a relapse in the first period, which is essential for survival, and chemotherapy for preventing death from state 1 (no relapse). Therefore, the probability of a relapse at any time in the first interval of time decreases when only radiotherapy is applied with respect to the no treatment case by 90.48%.

	$e^{[z-(1,1,1)'\boldsymbol{\beta}_{12}^{1}]}$			$e^{[z-(1,1,1)'\boldsymbol{\beta}_{13}^{1}]}$			$e^{[z-(1,1,1)'\boldsymbol{\beta}_{23}^{1}]}$		
z	I_1 ≡[0, 19)	I_2≡[19, 93)	I_3≡[93,∞)	I_1 ≡[0, 19)	I_2≡[19, 93)	I_3≡[93,∞)	I_1 ≡[0, 19)	I_2≡[19, 93)	I_3≡[93,∞)
RT	0.0952	0.3174	0.4366	10.404	0.7546	----	26.472	13.256	0.0053
HT	10.129	0.4918	12.525	19.061	0.8886	0.5206	0.4071	0.7656	0.0088
CT	0.4997	0.6962	0.2198	0.3029	0.7649	0.5610	20.913	11.200	320.9651
RT-HT	0.0964	0.1561	0.5468	19.832	0.6705	----	10.778	10.148	0.00005
RT-CT	0.0476	0.2210	0.0960	0.3151	0.5772	----	55.361	14.846	16.956
HT-CT	0.5061	0.3424	0.2753	0.5773	0.6797	0.2920	0.8515	0.8574	28.362
RT-HT-CT	0.0482	0.1087	0.1202	0.6006	0.5129	----	22.540	11.366	0.0150
No treatment	1	1	1	1	1	1	1	1	1

Table 4. Conversion factors with respect to no treatment for each interval of time and transitions 1→2, 1→3 and 2→3 according to the type of treatment.

Based on the piecewise Markov model, the estimated transition probability from no relapse to relapse depending on treatment is illustrated in Figure 3. These plots show that radiotherapy is essential in order to avoid local relapse. When radiotherapy is present, the probability of relapse decreases considerably.

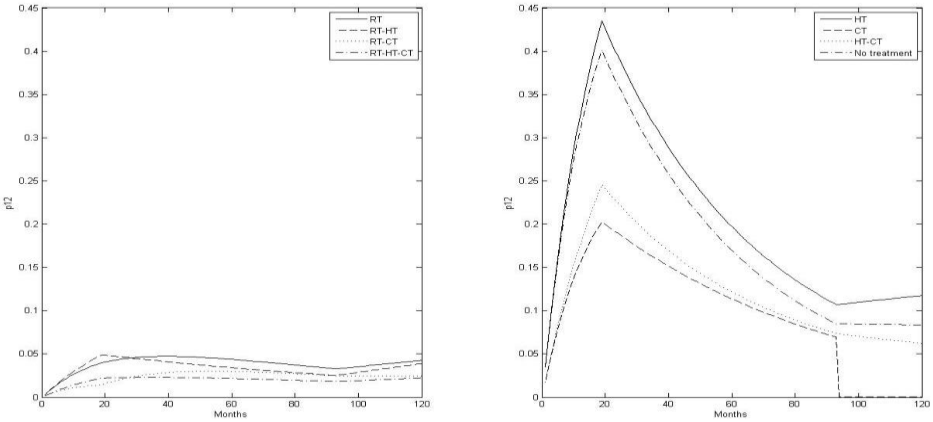

Figure 3. Estimated probability of being in the relapse state versus time according to the type of treatment

In Figures 4 and 5, the estimated survival functions for RT, RT-CT, RT-HT-CT and No treatment according to the piecewise Markov model are compared to the survival function by considering the estimated product limit. A period of 120 months is plotted, and the dashed lines are the confidence bands for the Kaplan-Meier estimate with a confidence level of 95%.

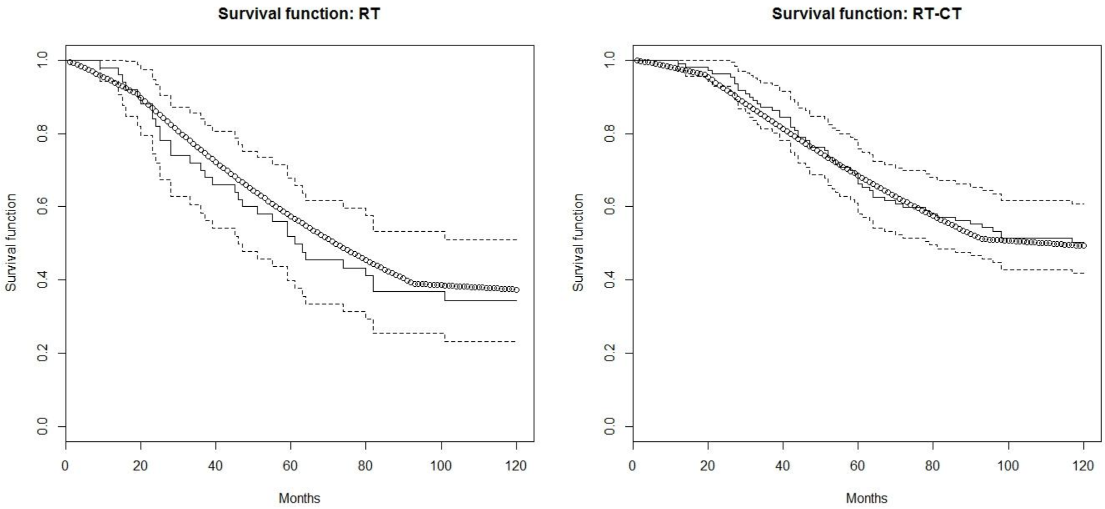

Figure 4. Estimated survival function for only RT and for RT-CT (circles) and the product limit (continuous line)

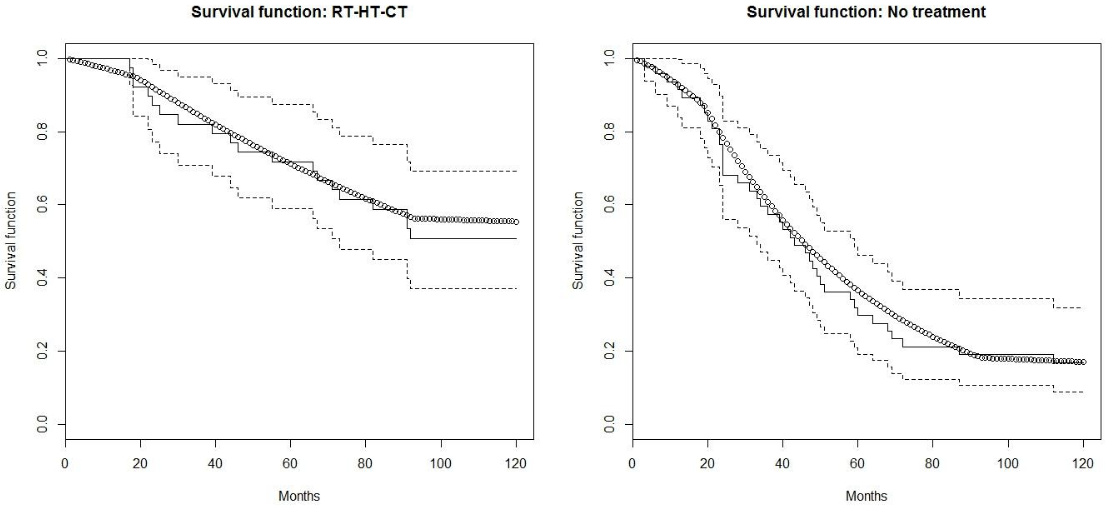

Figure 5. Estimated survival function for RT-HT-CT and for No treatment (circles) and the product limit (continuous line)

6.2.1 Goodness of fit

The goodness of fit was performed using several methods. First, the performances of the homogeneous model and the piecewise model were compared by using empirical estimates. Moreover, a Pearson-type goodness-of-fit test for the multi-state Markov model was applied to the homogeneous and piecewise models. Then, the goodness of fit of the piecewise multi-state model was evaluated using the Hollander and Proschan test (Hollander and Proschan, 1979).

Figures 4 and 5 show the 95% confidence intervals of the Kaplan-Meier estimates for RT and RT-CT and for RT-HT-CT and No treatment, respectively. The piecewise non-homogeneous model fits all cases better than the homogeneous model. Moreover, the estimated homogeneous survival curve from the fitted multi-state model for the RT-CT and No treatment group goes outside the confidence limits of the Kaplan-Meier estimates; this may be taken as informal evidence of a lack of fit. To measure the difference between the Kaplan-Meier curve and the estimated survival curves for a homogeneous and a piecewise model, we performed the following calculation:

$$\sum \left(S_{KM}(t) - S_0(t) \right)^2 ,$$

where $S_{KM}(t)$ is the value of the Kaplan-Meier estimate at time t and $S_0(t)$ is the survival function obtained from the model at time t. The results are reported in Table 5.

Treatments	Piecewise	Homogeneous	No. of patients
HT	0.20424	0.25178	7
RT	0.02970	0.27217	50
CT	0.40446	0.46479	13
HT-CT	0.14277	0.32417	12
RT-HT	0.05888	0.07382	22
RT-CT	0.18620	0.55041	110
RT-HT-CT	0.05715	0.23182	39
No treatment	0.04543	0.48377	47

Table 5. Results for the difference between the Kaplan-Meier curve and the estimated survival curve for a homogeneous and a piecewise model.

The results show a better fit for the piecewise model, as all of its values are smaller than the corresponding value from the homogeneous model.

Applying the Pearson-type goodness-of-fit test for the multi-state Markov model (Titman and Sharples, 2010), the value of the statistic was 481.888 for the homogeneous model and 29.914 for the piecewise model. The asymptotic null distribution is a χ^2 with 49 degrees of freedom[a] with $\chi^2(0.95) = 33.93$, indicating that the fit of the homogeneous model is poor, whereas the piecewise model seems to fit the data adequately. In Appendix B, we report the contingency tables for the observed and estimated counts for homogeneous (Table 11a) and piecewise models (Table 11b).

The second step in determining goodness of fit is to apply the Hollander and Proschan test for a multicensored data set. The experimental statistics are denoted by Z_{exp} and are normally distributed under the null hypothesis H_0: $S(t)=S_0(t)$, where S(t) is the underlying survival function and $S_0(t)$ is the survival function obtained from the piecewise model. Table 6 shows the results obtained by applying the test. In most cases, the model seems to fit adequately: the null hypothesis is accepted at a level of α=0.05. For chemotherapy and hormone-chemotherapy treatments, the null hypothesis can be rejected if a lower significance level is assumed.

[a] The degree of freedom is given by 7 possible transitions (1→1, 1→2, 1→3 1→C, 2→2, 2→3, 2→C), 3 periods, 8 groups of patients divided by treatment regimen and 35 estimated parameters: (7−1) x (3−1) x (8−1)−35=84−35=49.

Treatment	Z_{exp}	*p-value*	No. of patients
HT	–0.120	0.904	7
RT	1.410	0.159	50
CT	–2.418	0.0156	13
HT-CT	–2.524	0.012	12
RT-HT	–1.684	0.092	22
RT-CT	–1.669	0.0950	110
RT-HT-CT	–1.025	0.305	39
No treatment	1.072	0.170	47

Table 6. Values of the statistic Z_{exp} of the Hollander and Proschan test and the *p-value* for the survival fit to the treatments for the piecewise model.

6.3 Estimates of the model with treatments and endogenous factors (infected axillary glands and menopausal status) as covariates

Multiple endogenous factors associated with the patients were analysed. Some of them are specific to the patient, such as age and menopausal state, and others are prognostic factors, such as the number of infected axillary glands. Previous non-parametric and semi-parametric methods were used, and conclusions about these factors were reached. For each patient, the number of infected axillary glands and the menopausal state were observed, the second at the moment of diagnosis. The first factor was partitioned into three significant groups: one, two and more than two infected axillary glands. The mean times to death when the number of infected axillary glands for these three groups were 169.11 months, 97.33 months and 78.49 months, respectively (calculated from the product-limit estimator). Similarly, menopausal status was registered for each patient. Two groups were considered: pre- and perimenopause and postmenopause. Additionally, the mean time to death was determined from the product limit estimator. For the first group, it was equal to 140.83 months (130 patients, 62 censored), and for the second group, it was 96.25 months (170 patients, 60 censored). One of the main problems when more variables are considered is the number of patients in each subgroup. The mean time from the prognosis as a function of treatment and the number of infected axillary glands or menopausal status is shown in Table 7.

	RT	HT	CT	RT-HT	RT-CT	HT-CT	RT-HT-CT	No treatment
1 infected axillary gland	112.65 (35)	___ (3)	111.43 (7)	91.22 (9)	160.82 (38)	___ (2)	141.09 (11)	119.56 (16)
2 infected axillary glands	38.33 (12)	35.5 (4)	60 (4)	165.23 (10)	93.65 (59)	78.6 (5)	110.60 (20)	50.96 (23)
>2 infected axillary glands	29.67 (3)	___ (0)	46.50 (2)	26 (3)	117.26 (13)	41.20 (5)	118.13 (8)	34.5 (8)
Pre- and perimenopause	105.06 (11)	74.67 (3)	114.8 (5)	180.5 (6)	134.79 (51)	97 (6)	112.26 (27)	78.38 (21)
Post-menopause	86.23 (39)	76.75 (4)	63.5 (8)	85.81 (16)	106.8 (59)	58 (6)	119,07 (12)	57.04 (26)

Table 7. Mean time to death and number of patients (in brackets) according to treatment and number of infected glands and menopausal status

We assumed a discrete-time Markov model similar to the model developed in Section 6.2 but with two new covariates: menopausal status (ME) and number of infected axillary glands (INF). The menopausal status covariate takes a value of one when the patient is in a post-menopausal status and zero otherwise. The number of infected axillary glands covariate is partitioned into three different groups (1, 2 and more than 2 infected axillary glands). To include this covariate in the model, two dummy covariates are introduced. The baseline for the model is a patient with 1 infected axillary gland with a menopausal status before postmenopause and all treatments applied. The dummy covariates for the number of infected glands are 2 infected axillary glands and more than 2 infected axillary glands, which take the value of 1 when the number of infected glands is certain and 0 otherwise. The number of patients with one infected axillary gland was 121, while 137 patients had two infected axillary glands and 42 patients had more than two infected axillary glands.

The parameters were estimated with a maximum likelihood function; Table 8 shows these values (the order of the covariates is RT, HT, CT, Menopausal status, 2 infected axillary glands, and more than 2 infected axillary glands). For the last period of time, none of the patients had more than 2 infected axillary glands; therefore, this covariate was removed for this period of time.

Parameters	$\boldsymbol{I_1} \equiv$**[0, 19)**	$\boldsymbol{I_2}\equiv$**[19, 93)**	$\boldsymbol{I_3}\equiv$**[93,∞)**
λ_{12}^{l}	0.0008	0.0006	0.0001
(s.e.)	(0.0004)	(0.0003)	(0.0003)
λ_{13}^{l}	0.0008	0.0024	0.0000
(s.e.)	(0.0007)	(0.0008)	(0.0000)
λ_{23}^{l}	0.0281	0.0220	0.0000
(s.e.)	(0.0460)	(0.0111)	(0.0005)
$\boldsymbol{\beta}_{12}^{l}$	(2.2168, 0.0373, 0.8168, −0.1057, 1.1011, 1.3114)	(1.1308, 0.7214, 0.5033, 0.2746, 0.5952, 0.9183)	(0.9944, 0.1961, 1.3787, 0.9166, 0.6763, −11.2480)
(s.e.)	(0.4009, 0.3762, 0.3756, 0.3097, 0.4065, 0.4748)	(0.4080, 0.4911, 0.4146, 0.3836, 0.3957, 0.5955)	(1.3059, 1.2649, 1.4050, 1.1520, 1.2372)
$\boldsymbol{\beta}_{13}^{l}$	(−0.2613, −0.4467, 1.5089, −0.0958, 1.3061, 1.5179)	(0.2698, 0.2571, 0.5824, 0.1669, 1.0314, 1.2614)	(−13.7257, 1.4645, 1.5560, 1.0581, 3.1020, −10.2250)
(s.e.)	(0.6276, 0.5458, 0.5919, 0.5498, 0.6828, 0.8355)	(0.2861, 0.2425, 0.2462, 0.2169, 0.2428, 0.3443)	(1018.05, 1.3155, 1.1118, 0.8991, 1.2657, 653.776)
$\boldsymbol{\beta}_{23}^{l}$	(−0.7974, 0.8359, −0.7959, −0.0330, 0.5100, 0.0277)	(−0.3171, 0.4436, −0.2133, 0.7890, −0.1713, 0.3553)	(4.8182, 4.7228, −5.0444, −11.9300, 0.7900, 0)
(s.e.)	(1.0023, 1.2996, 0.9642, 0.8579,	(0.2948, 0.3953, 0.3111, 0.3051, 0.3390,	(15.0588, 15.0378, 15.1323, 947.881,

	1.1554, 1.5840)	0.4352)	1.4229, ---)

Table 8. Maximum likelihood estimates and standard errors of baseline transition probabilities and regression coefficients for the extended model

From the standard error, we can conclude that the covariate RT for the transition 1→3 and the covariate ME for the transition 2→3 can be removed for the last period of time.

The parameters were estimated for this reduced model, and these estimates and their standard errors for the last period I_3 are shown in Table 9.

Parameters	$I_3 \equiv [93,\infty)$
λ^{I}_{12}	0.0001
(s.e.)	(0.0002)
λ^{I}_{13}	0.0000
(s.e.)	(0.0001)
λ^{I}_{23}	0.0000
(s.e.)	(0.0004)
$\boldsymbol{\beta}^{I}_{12}$	(1.0401, –0.1305, 1.5219, 0.9035, 0.7933, 0)
(s.e.)	(1.2933, 1.2673, 1.3294, 1.2378, 1.3188, ---)
$\boldsymbol{\beta}^{I}_{13}$	(0, 1.2269, 1.3085, 0.9941, 3.0917, 0)
(s.e.)	(---, 1.2428, 1.0306, 0.8993, 1.1936, ---)
$\boldsymbol{\beta}^{I}_{23}$	(4.7713, 4.7297, –5.1266, 0, 0.9198, 0)
(s.e.)	(13.2052, 13.181, 13.2977, ---, 1.4386, ---)

Table 9. Maximum likelihood estimates and standard errors of baseline transition probabilities and regression coefficients in the third interval of time for the extended model

From these estimates, the estimated survival functions with respect to treatment, number of infected axillary glands and menopausal status are plotted in Figures 6-11, illustrating the effectiveness of treatments according to the different endogenous factors.

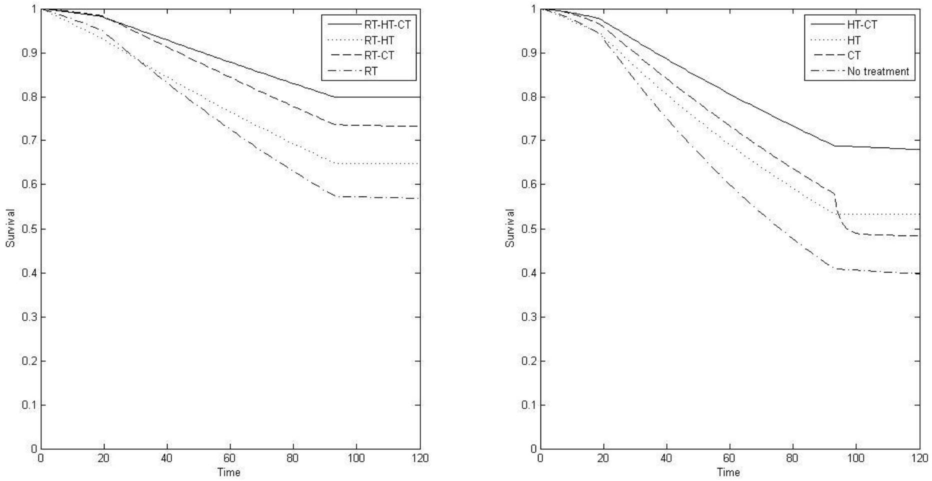

Figure 6. Estimated survival probability according to treatment for the one infected axillary gland and before post-menopause risk group

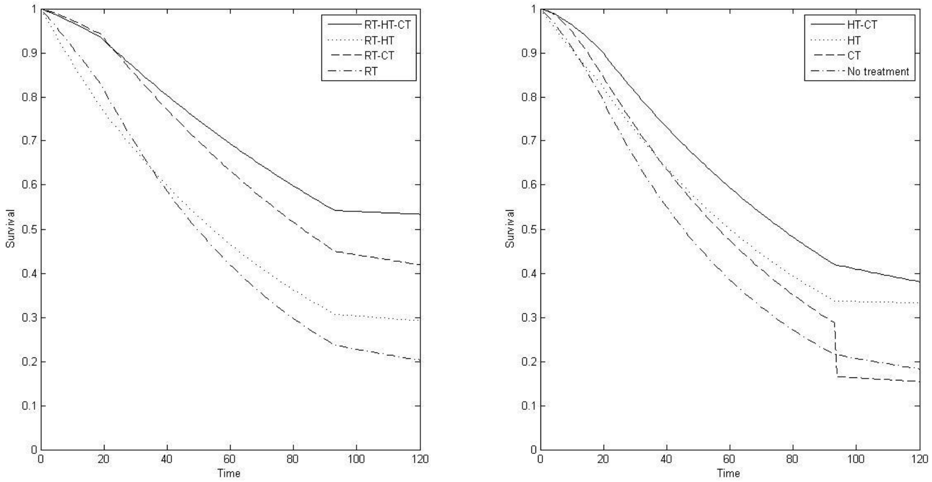

Figure 7. Estimated survival probability according to treatment for the two infected axillary glands and before post-menopause risk group

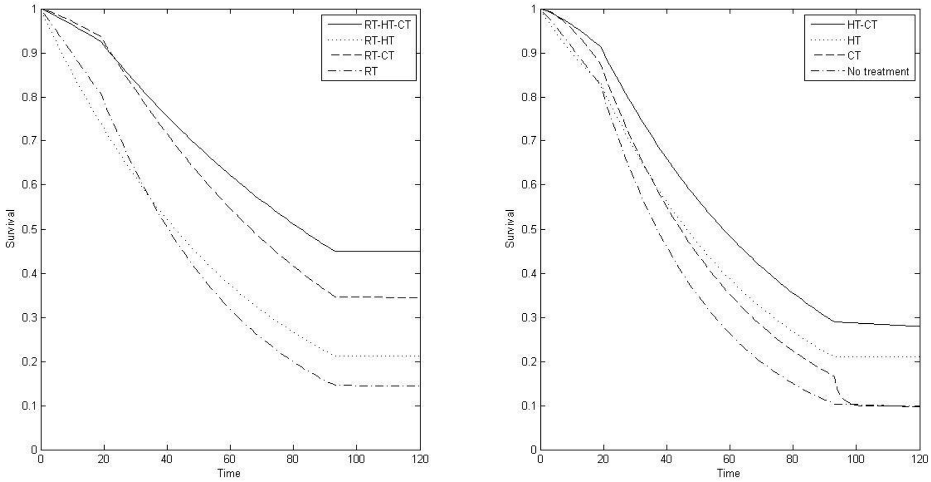

Figure 8. Estimated survival probability according to treatment for the more than two infected axillary glands and before post-menopause risk group

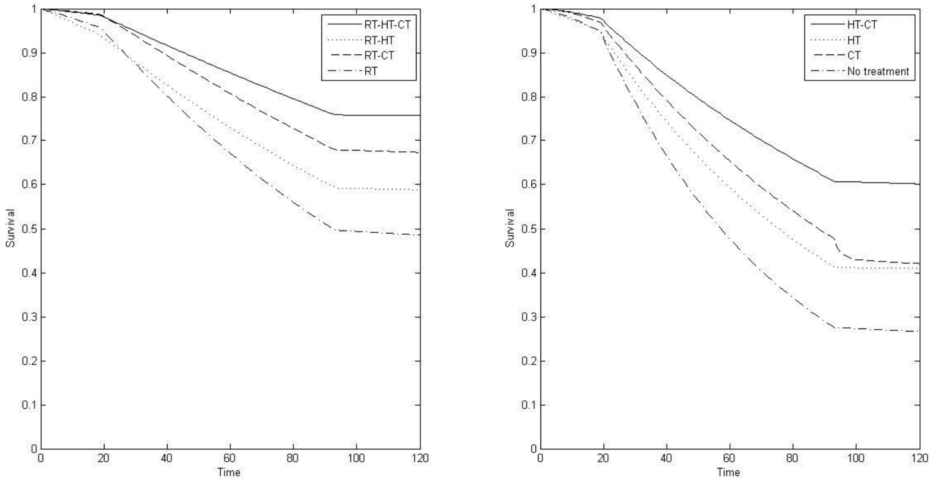

Figure 9. Estimated survival probability according to treatment for the one infected axillary gland and post-menopause risk group

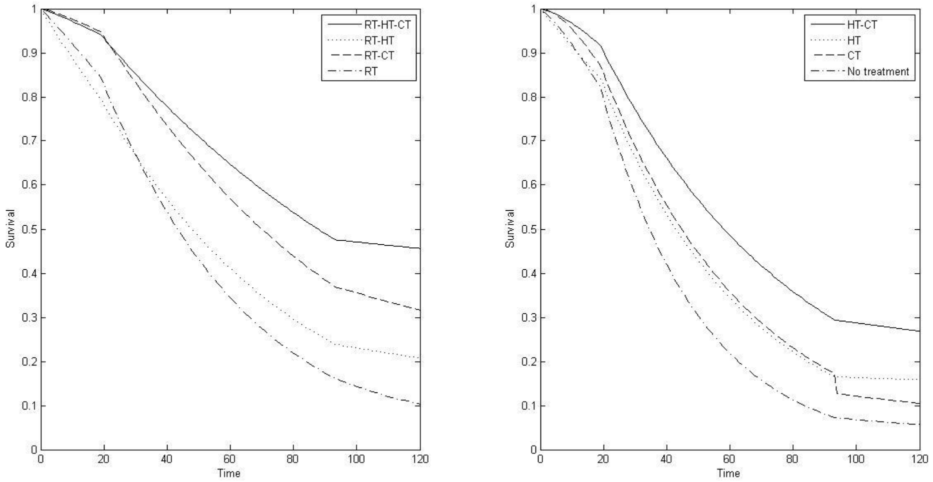

Figure 10. Estimated survival probability according to treatment for the two infected axillary glands and post-menopause risk group

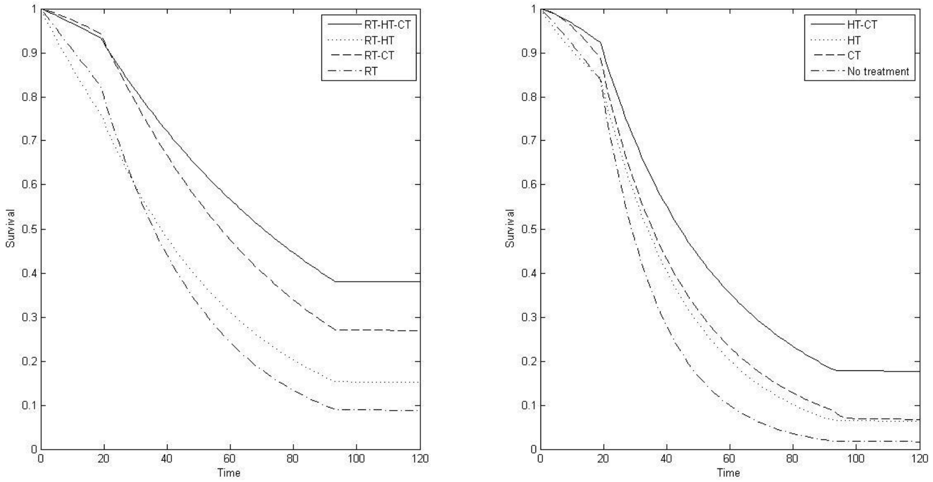

Figure 11. Estimated survival probability according to treatment for the more than two infected axillary glands and post-menopause risk group

6.4 Comparing the models

Here, we compare the models estimated in Sections 6.2 and 6.3. Table 10 reports the values of the log-likelihood, AIC and BIC for the models.

Model	**Log-likelihood**	**No. of parameters**	**AIC**	**BIC**
Without covariates	-1469.807	9	2957.613	2961.907
Full model (with only treatments)	-1419.979	36	2911.959	2929.135
Reduced model (with only treatments)	-1420.834	35	2911.668	2928.367
Full model (with treatments and endogenous factors)	-1385.346	66	2902.691	2934.181
Reduced model (with treatments and endogenous factors)	-1388.444	61	2898.888	2927.992

Table 10. Comparison among the estimated models.

As shown in Table 10, using both AIC and BIC values, the best model is the reduced model with treatments and endogenous factors.

7. Conclusions

In this paper, a general, discrete-time piecewise Markov process was developed to model the behaviour of a multi-state illness . The model and the likelihood function for two cases (cut-points known and unknown) were built in a matrix algorithmic form to facilitate their computational implementation. A first approximation of time-dependent covariate vectors was introduced, and the lifetime distribution and several important associated measures were worked out for several risk groups. This model was applied to analyse the progression of breast cancer from a data set of a cohort with 300 patients. The Markov hypothesis was tested; therefore, we assume that the Markov model is appropriate. This dataset was use by Pérez-Ocón et al. (1998, 2001) in order to analyse the evolution of breast cancer with continuous-time homogeneous and non-homogeneous Markov models.

A discrete-time homogeneous Markov model was considered initially to analyse the behaviour of this illness; however, the fits obtained for different risk groups were not sufficient. Consequently, the homogeneity of the data was tested, and we found no empirical evidence against the non-homogeneity of the data. An initial approximation of a non-homogeneous Markov model was developed as a piecewise model in this paper. Previous studies (investigating non-parametric and semi-parametric methods) led us to consider a model with two cut-points: 18 and 90 months. The likelihood function for estimating the parameters and the cut-points of the model was also built from the probability mass function. Two cut-points were estimated, 19 and 93 months. The remaining parameters were estimated, and the model and associated measures were subsequently determined.

The main analysis was developed to study the evolution of the illness using the treatments as covariate, similar to the models developed in continuous cases. The effect of treatments on the evolution of the survival probabilities and the relapse times were determined. Relapse is very important in the progression of breast cancer, and here it has been examined in detail with respect to different risk groups associated with treatment. The empirical survival function was compared with the estimated survival functions for the homogeneous and piecewise non-homogeneous Markov models and different risk groups.

A goodness-of-fit study was carried out, which proved that the discrete piecewise model improves the discrete homogeneous model. Furthermore, the fit of this discrete model improves upon those of the continuous-time cases; both the homogeneous case (Pérez-Ocón et al., 1998) and the non-homogeneous case with a cut-point (Pérez-Ocón et al. (2001). Thus, the expected survival time and other measures from the discrete-time non-homogeneous model with two cut-point are more reliable than the others described. Conditional probabilities of relapse and death were estimated and shown at certain units of time.

Finally, these models were extended by introducing endogenous factors through covariates. This extension was not included in previous works. Multiple factors were analysed through non-parametric and semi-parametric methods. We observed that age was not a significant factor if the patient’s menopausal status was available, while the number of infected axillary glands was a significant factor. A new discrete model with treatments, menopausal status and number of infected axillary glands was built through dummy covariates. The effectiveness of the treatments was shown from the estimated survival probability according to different risk groups (different numbers of infected axillary glands and menopausal status). As expected, the risk increased with the number of infected axillary gland.

The results were implemented computationally using the MATLAB and R programs.

Acknowledgements

Funding: This paper is partially supported by the Junta de Andalucía, Spain under grant FQM-307, by the Ministerio de Economía y Competitividad, Spain under grant MTM2017-88708-P, by the European Regional Development Fund (ERDF), and by the University of Milano-Bicocca under grant number 2014-ATE-0228.

APPENDIX A

The parameters of the model are estimated by a maximum likelihood function. These parameters are the matrices $\mathbf{T}_u$ (or parameters inside these matrices), the regression covariate vectors $\boldsymbol{\beta}^u$, for $u = 1,...,k$ and the cut-points, all of them estimated jointly. We assume that n items are observed, all beginning in state 1, and item i is observed at m_i change times, the last time being death or censorship. Given that the item is observed at change times, then for any item, the value of the covariate vector and the corresponding state is observed. Therefore, a sequence of times, states and values of the covariate vector is achieved for each item i: $0 = t_{i,1} < t_{i,2} < \cdots < t_{i,m_i}$, $1 = x^i_1, \ldots, x^i_{m_i}$ and $\mathbf{z}^i_{l_1}, \ldots, \mathbf{z}^i_{l_{m_i}}$, respectively. $\mathbf{z}^i_{l_s}$ corresponds to the covariate vector for the interval that contains the time $t_{i,s}$ for item i and for $s = 1,...,m_i$.

We assume k–1 unknown positive integer cut-points, c_0=0 <c$_1$ <…<c_{k-1}<c_k=∞. The likelihood function for estimating the parameters is given by

$$L\left(c_1,\ldots,c_{k-1},\mathbf{T}_u,\boldsymbol{\beta}^u,u=1,...,k\right)=\prod_{i=1}^{n}\prod_{s=2}^{m_i} h_{x^i_{s-1},x^i_s}\left(\mathbf{T}_u,\boldsymbol{\beta}^u,u=1,...,k \,\middle|\, t_{i,s-1},t_{i,s},\mathbf{z}^i_{l_{s-1}},\ldots,\mathbf{z}^i_{l_s}\right).$$

For the calculations, we define the intervals $I_q = \left[c_{q-1},c_q\right[\,; J_q = \left]c_{q-1},c_q\right],\ j=1,...,k$. Let $f^q_x\left(t,\mathbf{z}^i_q;\mathbf{T}_q,\boldsymbol{\beta}^q\right)$ be the sojourn time probability in state x at time t calculated by using the matrix $\mathbf{P}_q\left(\mathbf{z}^i_q\right)$. Given that the state at any cut-point is known, then the factors in the likelihood function have the following expressions,

i) If $t_{i,s-1}$ and $t_{i,s}$ belong to intervals I_j and J_j, respectively,

$$h_{x^i_{s-1},x^i_s}\left(\mathbf{T}_j,\boldsymbol{\beta}^j \,\middle|\, t_{i,s-1},t_{i,s},\mathbf{z}^i_{l_{s-1}},\ldots,\mathbf{z}^i_{l_s}\right) = f^j_{x^i_{s-1}}\left(t_{i,s}-t_{i,s-1}-1,\mathbf{z}^i_j;\mathbf{T}_j,\boldsymbol{\beta}^j\right)T^j_{x^i_{s-1}x^i_s}\left(\mathbf{z}^i_j\right).$$

ii) If $t_{i,s-1}$ and $t_{i,s}$ belong to interval I_{j-1}, J_j, respectively,

$$\begin{aligned} h_{x^i_{s-1},x^i_s}\left(\mathbf{T}_u,\boldsymbol{\beta}^u,u=j-1,j \,\middle|\, t_{i,s-1},t_{i,s},\mathbf{z}^i_{l_{s-1}},\ldots,\mathbf{z}^i_{l_s}\right) &= f^{j-1}_{x^i_{s-1}}\left(c_{j-1}-t_{i,s-1},\mathbf{z}^i_{j-1};\mathbf{T}_{j-1},\boldsymbol{\beta}^{j-1}\right) \\ &\cdot f^j_{x^i_{s-1}}\left(t_{i,s}-c_{j-1}-1,\mathbf{z}^i_j;\mathbf{T}_j,\boldsymbol{\beta}^j\right)T^j_{x^i_{s-1},x^i_s}\left(\mathbf{z}^i_j\right). \end{aligned}$$

iii) If $t_{i,s-1} \in I_j$ and $t_{i,s} \in J_q$ with $q - j \geq 2$,

$$\begin{aligned} h_{x^i_{s-1},x^i_s}\left(\mathbf{T}_u,\boldsymbol{\beta}^u,u=j,...,q \,\middle|\, t_{i,s-1},t_{i,s},\mathbf{z}^i_{l_{s-1}},\ldots,\mathbf{z}^i_{l_s}\right) &= f^j_{x^i_{s-1}}\left(c_j-t_{i,s-1},\mathbf{z}^i_j;\mathbf{T}_j,\boldsymbol{\beta}^j\right) \\ &x\prod_{u=j+1}^{q-1} f^u_{x^i_{s-1}}\left(c_u-c_{u-1},\mathbf{z}^i_u;\mathbf{T}_u,\boldsymbol{\beta}^u\right)f^q_{x^i_{s-1}}\left(t_{i,s}-c_q-1,\mathbf{z}^i_q;\mathbf{T}_q,\boldsymbol{\beta}^q\right)T^q_{x^i_{s-1},x^i_s}\left(\mathbf{z}^i_q\right). \end{aligned}$$

The likelihood function is maximized by considering several restrictions. The matrices $\mathbf{P}_q$ and $\mathbf{P}_q\left(\mathbf{z}_q^i\right)$ associated with the model should be stochastic matrices for any covariate vector $\mathbf{z}_q^i$. This restriction will not allow probabilities less than zero or greater than one for any values of the parameters.

Then, the cut-points are estimated, and the optimum values $c_1,\ldots,c_{k-1}$ are the values that verify

$$c_1,\ldots,c_{k-1} \in \mathrm{N} \text{ such that } L\left(c_1,\ldots,c_{k-1},\hat{\mathbf{T}}_u^{c_1,\ldots,c_{k-1}},\hat{\boldsymbol{\beta}}_u^{c_1,\ldots,c_{k-1}},u=1,...,k\right)=\max_{v_j}\left\{L\left(v_1,\ldots,v_{k-1},\hat{\mathbf{T}}_u^{v_1,\ldots,v_{k-1}},\hat{\boldsymbol{\beta}}_u^{v_1,\ldots,v_{k-1}},u=1,...,k\right)\right\},$$

subject to $0<v_j<v_{j+1}$ for $j=1,...,k-2$ and $v_{k-1}<\max_i\left\{t_{i,m_i}\right\}$, where v_j belongs to the set of natural numbers for any j with the corresponding restrictions. $\left(\hat{\mathbf{T}}_u^{v_1,\ldots,v_{k-1}},\hat{\boldsymbol{\beta}}_u^{v_1,\ldots,v_{k-1}},u=1,...,k\right)$ are the maximum likelihood estimates of $\left(\mathbf{T}^u,\boldsymbol{\beta}^u,u=1,...,k\right)$ for $v_1,\ldots,v_{k-1}$.

The likelihood function has been implemented computationally with Matlab and it is maximized by using the function *fmincon* of this programme. This function is used to find the minimum of a constrained nonlinear multivariable function by using the interior-point algorithm.

APPENDIX B

First period: [0,19]	**RT**	**Observed**	42	0	3	0	5	0	0	0
		Expected	42.0006	0	1.8601	0	6.1393	0	0	0
	HT	**Observed**	5	0	2	0	0	0	0	0
		Expected	4.7696	0	1.2585	0	0.9719	0	0	0
	CT	**Observed**	10	0	2	0	1	0	0	0
		Expected	9.9059	0	1.5631	0	1.5310	0	0	0
	RT-HT	**Observed**	18	0	1	0	3	0	0	0
		Expected	18.7979	0	0.6549	0	2.5472	0	0	0
	RT-CT	**Observed**	106	0	2	0	2	0	0	0
		Expected	98.6228	0	2.1036	0	9.2736	0	0	0
	HT-CT	**Observed**	6	0	5	0	1	0	0	0
		Expected	9.6254	0	1.1872	0	1.1874	0	0	0
	RT-HT-CT	**Observed**	35	0	1	0	3	0	0	0
		Expected	35.2960	0	0.6064	0	3.0976	0	0	0
	No treatment	**Observed**	21	0	19	0	7	0	0	0
		Expected	29.0528	0	10.2311	0	7.7161	0	0	0
Second period: (19,93]	**RT**	**Observed**	16	3	0	0	23	0	0	3
		Expected	17.9361	3.3630	2.4381	0	18.2628	0.4273	0	25727
	HT	**Observed**	3	0	0	0	2	0	0	2
		Expected	1.1221	0	1.3578	0	2.5201	0.6838	0	13162

	CT	**Observed**	5	0	0	0	5	0	0	2
		Expected	3.4692	0	1.4724	0	5.0585	0.2246	0	17754
	RT-HT	**Observed**	7	4	1	0	6	0	0	1
		Expected	6.2075	3.5472	1.0753	0	7.1702	0.2986	0	0.7014
	RT-CT	**Observed**	53	3	4	0	46	0	0	2
		Expected	65.5726	3.7117	3.1534	0	33.5623	0.1700	0	18300
	HT-CT	**Observed**	2	0	1	0	3	0	0	5
		Expected	2.5420	0	1.0251	0	2.4329	1.2892	0	37108
	RT-HT-CT	**Observed**	18	1	1	0	15	0	0	1
		Expected	22.4797	1.2489	1.0934	0	10.1781	0.2172	0	0.7828
	No treatment	**Observed**	4	0	2	0	15	3	0	16
		Expected	3.2253	0	4.6706	0	13.1041	3.3701	0	156299
Third period (93, last time observed]	**RT**	**Observed**	0	14	0	1	1	0	0	0
		Expected	0	7.1358	0	0.8850	7.9792	0	0	0
	HT	**Observed**	0	3	0	0	0	0	0	0
		Expected	0	1.3377	0	0.7901	0.8722	0	0	0
	CT	**Observed**	0	5	0	0	0	0	0	0
		Expected	0	1.8632	0	0.7608	2.3760	0	0	0
	RT-HT	**Observed**	0	6	0	0	1	0	1	0
		Expected	0	1.9400	0	0.3396	4.7203	0	0.0795	0.9205
	RT-CT	**Observed**	0	49	0	0	4	0	3	1

		Expected	0	29.3247	0	1.4445	22.2308	0	0.1294	3.8706
	HT-CT	**Observed**	0	2	0	0	0	0	0	1
		Expected	0	0.5389	0	0.2951	1.1660	0	0.1262	0.8738
	RT-HT-CT	**Observed**	0	17	0	1	0	0	1	0
		Expected	0	10.5343	0	0.5655	6.9002	0	0.1219	0.8781
	No treatment	**Observed**	0	3	0	1	0	0	4	1
		Expected	0	0.0714	0	0.2421	3.6864	0	0.1216	4.8784

Table 11a. Contingency table of observed and expected counts for the homogeneous model.

			1→1	**1→1C**	**1→2**	**1→2C**	**1→3**	**2→2**	**2→2C**	**2→3**
First period: [0,19]	**RT**	**Observed**	42	0	3	0	5	0	0	0
		Expected	43.3841	0	1.9887	0	4.6271	0	0	0
	HT	**Observed**	5	0	2	0	0	0	0	0
		Expected	3.0346	0	3.0446	0	0.9208	0	0	0
	CT	**Observed**	10	0	2	0	1	0	0	0
		Expected	9.0789	0	2.6324	0	1.2888	0	0	0
	RT-HT	**Observed**	18	0	1	0	3	0	0	0
		Expected	17.7601	0	1.0785	0	3.1614	0	0	0
	RT-CT	**Observed**	106	0	2	0	2	0	0	0
		Expected	104.0731	0	1.5761	0	4.3508	0	0	0
	HT-CT	**Observed**	6	0	5	0	1	0	0	0
		Expected	8.1708	0	2.9521	0	0.8771	0	0	0
	RT-HT-CT	**Observed**	35	0	1	0	3	0	0	0

		Expected	36.0999	0	0.8595	0	2.0407	0	0	0
	No treatment	**Observed**	21	0	19	0	7	0	0	0
		Expected	22.067	0	18.8432	0	6.0898	0	0	0
Second period: (19.93]	**RT**	**Observed**	16	3	0	0	23	0	0	3
		Expected	14.5674	2.7314	1.5034	0	23.1978	0.131	0	2.869
	HT	**Observed**	3	0	0	0	2	0	0	2
		Expected	1.622	0	0.3973	0	2.9807	0.3332	0	1.6668
	CT	**Observed**	5	0	0	0	5	0	0	2
		Expected	3.1561	0	0.7861	0	6.0579	0.1432	0	1.8568
	RT-HT	**Observed**	7	4	1	0	6	0	0	1
		Expected	5.6735	3.242	0.4599	0	8.6246	0.0921	0	0.9079
	RT-CT	**Observed**	53	3	4	0	46	0	0	2
		Expected	51.705	2.9267	2.7436	0	48.6246	0.0595	0	1.9405
	HT-CT	**Observed**	2	0	1	0	3	0	0	5
		Expected	2.5988	0	0.3572	0	3.044	0.67	0	4.33
	RT-HT-CT	**Observed**	18	1	1	0	15	0	0	1
		Expected	19.525	1.0847	0.6298	0	13.7605	0.0688	0	0.9312
	No treatment	**Observed**	4	0	2	0	15	3	0	16
		Expected	4.3573	0	2.0811	0	14.5616	1.8127	0	17.1873
Third period (93. last time observed]	**RT**	**Observed**	0	14	0	1	1	0	0	0
		Expected	0	11.9575	0	1.2539	2.7886	0	0	0
	HT	**Observed**	0	3	0	0	0	0	0	0
		Expected	0	2.6664	0	0.3334	0.0002	0	0	0

	CT	**Observed**	0	5	0	0	0	0	0	0
		Expected	0	4.8249	0	0	0	0	0	0
	RT-HT	**Observed**	0	6	0	0	1	0	1	0
		Expected	0	4.7707	0	1.1591	1.0702	0	1	0
	RT-CT	**Observed**	0	49	0	0	4	0	3	1
		Expected	0	45.3907	0	0.8756	6.7337	0	2.3183	1.6817
	HT-CT	**Observed**	0	2	0	0	0	0	0	1
		Expected	0	1.859	0	0.0889	0.0521	0	0.3669	0.6331
	RT-HT-CT	**Observed**	0	17	0	1	0	0	1	0
		Expected	0	16.3407	0	0.4932	1.1661	0	0.9953	0.0047
	No treatment	**Observed**	0	3	0	1	0	0	4	1
		Expected	0	2.752	0	0.9703	0.2777	0	3.0447	1.9553

Table 11b. Contingency table of observed and expected counts for the piecewise model.

References

1. Andersen, P.K. and Keiding, N. (2001) Multi-state models for event history analysis. *Statistical Methods in Medical Research*, **11**, 91-115.
2. Bacchetti, P.; Boylan, R.D.; Terrault, N.A.; Monto, A. and Berenguer, M. (2010) Non-Markov multistate modeling using time-varying covariates, with application to progression of liver fibrosis due to hepatitis C following liver transplant. *The International Journal of Biostatistics*, **6**, 1, 1-14.
3. Chen, B., Yi, G.Y. and Cook, R.J. (2010) Analysis of interval censored disease progression data via multistate models under a non ignorable inspection process. *Statistics in Medicine*, **29**, 1175–1189.
4. Commenges, D. and Joly, P. (2001) Multi-state model for dementia, institutionalization and death. *Communications in Statistics*-A, **33**, 1315–1326.
5. Cortese, G. and Andersen, P.K. (2010) Competing Risks and Time-Dependent Covariates. *Biometrical Journal*, **52**, 1, 138–158.
6. Faddy, M.J. (1998) On inferring the number of phases in a coxian phase-type distribution. *Communications in Statistics. Stochastic Models*, **14**,1-2, 407-417
7. Farewell, V.T., Tom, B.D.M. (2014) The versatility of multi-state models for the analysis of longitudinal data with unobservable features. *Lifetime data analysis*, **20**, 51-75.
8. Hollander, M. and Proschan, F. (1979) Testing to Determine the Underlying Distribution Using Randomly Censored Data. *Biometrics*, **35**, 2, 393-401.
9. Hougaard, P. (1999) Multi-state models: a review. *Lifetime Data Analysis*, **5**, 239-264.
10. Ieva, F., Jackson, C. and Sharples, L.D. (2015) Multi-State modelling of repeated hospitalisation and death in patients with Heart Failure: the use of large administrative databases in clinical epidemiology. *Statistical Methods in Medical Research*, doi: 10.1177/0962280215578777
11. Jackson, C.H. (2011) Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*, 38: 1–29.
12. Jackson, C.H., Sharples, L.D., Thompson, S.G. and Duffy, S.W., Couto, E. (2003) Multi-state Markov models for disease progression with classification error. *Statistician*, **52**, 193–209.
13. Kalbfleisch, J.D. and Lawless, J.F. (1985) The Analysis of Panel Data Under a Markov Assumption. *Journal of the American Statistical Association*, **80**, 863-871.
14. Kalbfleisch, J.D. and Prentice, R.L. (1980) *The Statistical Analysis of Failure Time Data*. Wiley Series in Probability and Mathematical Statistics.
15. Meira-Machado, L., de Uña-Alvarez, J. and Cadarso-Suarez, C. (2009) Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research*, **18**,2,195-222.

16. Neuts, M.F. (1981) *Matrix-geometric solutions in stochastic models.* Volume 2 of Johns Hopkins Series in the Mathematical Sciences.
17. Pérez-Ocón, R., Ruiz-Castro J.E. and Gámiz-Pérez, M.L. (1998) A Multivariate Model to Measure the Effect of Treatments in Survival to Breast Cancer *Biometrical Journal*, **40**, 6, 703-715.
18. Pérez-Ocón, R., Ruiz-Castro, J.E. and Gámiz-Pérez M.L. (2001) Non-homogeneous Markov Processes for Analysing the Effect of Treatments to Breast Cancer. *Statistics in Medicine*, **20**, 109-122.
19. Putter, H., Fiocco, M. and Geskus, R.B. (2007) Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, **26**, 2389-2430.
20. Santamaría, C., García-Mora, B., Rubio, G. and Navarro, E. (2009)A Markov model for analyzing the evolution of bladder carcinoma. *Mathematical and Computer Modelling*, **50**, 726-732.
21. Singer, J.D. and Willett, J.B. (2003) *Applied Longitudinal Data Analysis*. Oxford: Oxford University Press.
22. Titman, A.C. (2014) Estimating parametric semi-Markov models from panel data using phase-type approximations. *Statistics and Computing*, **24**, 155–164.
23. Van De Hout, A. (2016) *Multi-state survival models for interval-censored data.* Boca Raton: CRC Press.