Journal of Modern Applied Statistical Methods

Volume 8 | Issue 2 Article 34

11-2009

Markov Modeling of Breast Cancer

Chunling Cong
University of South Florida

Chris P. Tsokos
University of South Florida

DOI: 10.22237/jmasm/1257035580

Journal of Modern Applied Statistical Methods
November 2009, Vol. 8, No. 2, 626-631

1538 – 9472/09/$95.00

Markov Modeling of Breast Cancer

Chunling Cong Chris P. Tsokos
University of South Florida

Previous work with respect to the treatments and relapse time for breast cancer patients is extended by applying a Markov chain to model three different types of breast cancer patients: alive without ever having relapse, alive with relapse, and deceased. It is shown that combined treatment of tamoxifen and radiation is more effective than single treatment of tamoxifen in preventing the recurrence of breast cancer. However, if the patient has already relapsed from breast cancer, single treatment of tamoxifen would be more appropriate with respect to survival time after relapse. Transition probabilities between three stages during different time periods, 2-year, 4-year, 5-year, and 10-year, are also calculated to provide information on how likely one stage moves to another stage within a specific time period.

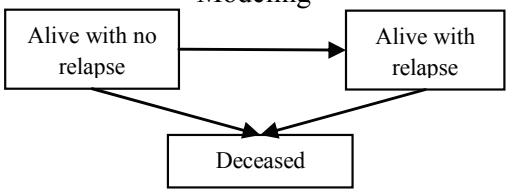
Key words: Markov chain, breast cancer, relapse time, tamoxifen and radiation.

Introduction

The Markov (1906) chain model has been applied in various fields such as physics, queuing theory, internet application, economics, finance, and social sciences among others. As an effective and efficient way of describing a process in which an individual moves through a series of states (stages) in continuous time, homogeneous Markov models have also been extensively used in health sciences where the progression of certain diseases are of great importance to both doctors and patients. In the present study, the main objective is to investigate the progression of breast cancer in patients in three different stages who were given different treatments. One group of patients received combined treatments of tamoxifen and radiation, and the other group received only tamoxifen. Figure 1 shows the three stages of interest in the study are: alive with no relapse, alive with relapse, and deceased. Even though breast cancer patients who have recurrence may be treated and recover from breast cancer to become active with no relapse, due to the fact that the data does not include any observations of that process, we consider the second state-alive with relapse as those patients who once had relapse and are still alive, regardless of whether they have recovered from breast cancer or not.

Chunling Cong is a doctoral student in Statistics at the University of South Florida. Her research interests are in developing forecasting and statistical analysis and modeling of cancer. Email: ccong@mail.usf.edu. Chris Tsokos is a Distinguished University Professor in mathematics and Statistics at the University of South Florida. His research interests are in modeling Global Warming, analysis and modeling of cancer data, parametric, Bayesian and nonparametric reliability, and stochastic systems, among others. He is a fellow of both ASA and ISI. Email: profcpt@cas.usf.edu.

Figure 1: Three Stages of Breast Cancer Modeling

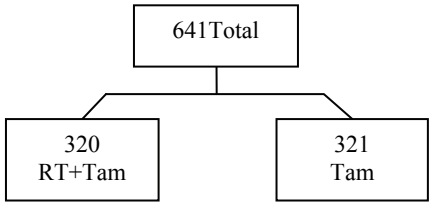


Methodology

Between December 1992 and June 2000, a total of 769 women were enrolled and randomized in the study. Among these, 386 received combined radiation and tamoxifen (RT+Tam), and the

remaining 383 received tamoxifen (Tam) only. The last follow-up was conducted in the summer of 2002. As shown in Figure 2, only those 641 patients enrolled at the Princess Margaret Hospital are included: 320 and 321 in RT+Tam and Tam treatment groups, respectively.

Figure 2: Breast Cancer Data



This data was used by Fyles, et al. and was later analyzed by Ibrahim, et al. Analysis was conducted on this data with respect to the treatment effect of the two different treatments using decision tree and modeled relapse time using AFT and Cox-PH model. Mixture models were also applied to compare the cure rate of the two groups.

The Markov Chain Model

The Markov chain is a model for a finite or infinite random process sequence $X = \{X_1, X_{2,...,X_N}\}$. Unlike the independent identical distribution (i.i.d) model that assumes the independency of a sequence of events X_i's, the Markov model takes into account the dependencies among the X_i's.

Consider a random process $X = \{X_t\}_{t\geq 1} = \{X_1, X_2, ...\}$ of random variables taking values in a discrete set space of stages $S = \{1, 2, 3, ..., s\}$ where X_t represents the state of the process of an individual at time *t*. The transitions possible among the three stages in this study, alive without relapse, alive with relapse, and deceased are shown in Figure 1 indicated by arrows. Consider a realization of the history of the process up to and including time *t*, as $\{X_t = x_t, X_{t-1} = x_{t-1}, ..., X_1 = x_1\}$, where $x_t, x_{t-1}, ..., x_1$ is a sequence of stages at different times. A random process is called a Markov Chain if the conditional probabilities between the stages at different times satisfy the Markov property: the conditional probability of future one-step-event conditioned on the entire past of the process is just conditioned on the present stage of the process. In other words, the one-step future stage depends only on the present stage:

$$P(X_{t+1} = x_{t+1} \mid X_t = x_t, X_{t-1} = x_{t-1}, ..., X_1 = x_1) \quad (1)$$
$$= P(X_{t+1} = x_{t+1} \mid X_t = x_t)$$

for every sequence $x_1, ..., x_t, x_{t+1}$ of elements of S and every $t \geq 1$.

The transition probability from stage i to stage j at time t and transition intensity are defined by

$$p_{ij}(t) = p(X_{t+1} = j \mid X_t = i), \quad (2)$$

and

$$q_{ij}(t) = \lim_{h\to 0} \frac{P(X(t+h) = j \mid X(t) = i)}{h}, \quad (3)$$

where h is the time interval.

If the transition probabilities do not depend on time, $p_{ij}(t)$ can simply be written as p_{ij}, then the Markov chain is called time-homogeneous. If not specified, the following analysis is based on time-homogeneous Markov chain. A transition probability matrix $P(t)$ consisting of all the transition probabilities between stages in a matrix form is given by:

$$P(t) = \begin{Bmatrix} p_{11}(t) & p_{12}(t) & ... & p_{1s}(t) \\ p_{21}(t) & p_{22}(t) & ... & p_{2s}(t) \\ ... & ... & ... & ... \\ p_{s1}(t) & p_{s2}(t) & ... & p_{ss}(t) \end{Bmatrix}, \quad (4)$$

where probabilities in each row add up to 1. Thus, it is 100% certain that for any individual at time *t* is in one of the stages and the sum of probabilities of being in each stage is 1.

The transition probability matrix can be calculated by taking the matrix exponential of the scaled transition intensity matrix defined by

$$P(t) = Exp(tQ), \quad (5)$$

where

$$Q = \begin{Bmatrix} q_{11} & q_{12} & \cdots & q_{1s} \\ q_{21} & q_{22} & \cdots & q_{2s} \\ \cdots & \cdots & \cdots & \cdots \\ q_{s1} & q_{s2} & \cdots & q_{ss} \end{Bmatrix}, \quad (6)$$

and q_{ij} denotes the transition intensity from stage i to stage j.

The exponential of a matrix A is defined by

$$Exp(A) = 1 + A^2/2! + A^3/3! + \ldots, \quad (7)$$

where each summand in the series is the matrix products. In this manner, once the intensity matrix is given, the transition probabilities can be calculated as shown above.

Next, the intensity matrix and transition probabilities matrix can be obtained by maximizing the likelihood $L(Q)$ which is a function of Q. Consider an individual consisting of a series of times $(t_1, t_2, ..., t_n)$ and corresponding stages $(x_1, x_2, ..., x_n)$. More specifically, consider a pair of successive stages observed to be i and j at time t_i and t_j. Three scenarios are proposed and considered here.

Scenario 1

If the information for the individual is obtained at arbitrary observation times (the exact time of the transition of stages is unknown) the contribution to the likelihood from this pair of states is:

$$L_{ij} = p_{ij}(t_j - t_i). \quad (8)$$

Scenario 2

If the exact times of transitions between different stages are recorded and there is no transition between the observation times, the contribution to the likelihood from this pair of stages is:

$$L_{ij} = p_{ij}(t_j - t_i) q_{ij}. \quad (9)$$

Scenario 3

If the time of death is known or $j = death$, but the stage on the previous instant before death is unknown as denoted by k (k could be any possible stage between stage i and death), the contribution to the likelihood function from this pair of stages is:

$$L_{ij} = \sum_{k \neq j} p_{ik}(t_j - t_i) q_{kj}. \quad (10)$$

Results

The breast cancer patients were divided into two groups RT+Tam and Tam based on the different treatments they received. For those patients who received combined treatments, 26 patients experienced relapse, 13 patients died without recurrence of breast cancer during the entire period of the study, and 14 died after recurrence of breast cancer. For the patients in the Tam group, 51 patients experienced relapse, 10 died without reoccurrence of breast cancer, and 13 died after recurrence of breast cancer.

As can be observed from the transition intensity matrixes for both groups RT+Tam and Tam as shown in Tables 1 and 2, patients who received single treatment have a higher transition intensity form Stage 1 to Stage 2, thus, they are more likely to have breast cancer recurrence. Thus, the probability of that happening in the Tam group is higher than that of the RT+Tam group. For those patients who died without relapse, there is no significant difference between the two treatments as illustrated by the intensity form Stage 1 to Stage 3.

Combined treatment is also more effective than a single treatment with respect to the possibility of death without relapse as can be observed from the transition intensity from Stage 1 to Stage 3. However, for those who already experienced relapse of breast cancer, patients who received combined treatments are more likely to die than those who received a single treatment. Therefore, combined treatment should be chosen over single treatment to avoid recurrence, but for those patients who already had breast cancer relapse, it would be advisable to choose a single treatment to extend the time from recurrence to death.

Figures 3 and 4 illustrate the effectiveness of the two treatments with respect to the survival probabilities and also show the survival curves of the patients who had recurrence and who had no recurrence in each treatment group.

From the above analysis, the proposed Markov chain model provides recommendations for which treatment to choose for breast cancer patients with respect to relapse and survival time. Moreover, it provides patients with very important information on the exact time or possibilities of recurrence and death. Estimated mean sojourn times in each transient stage for patients who received combined treatment are 43.46 and 3.25 in Stage 1 and Stage 2, respectively. Estimated mean sojourn times for patients who received single treatment are 25.53 and 11.72 in Stage 1 and Stage 2. This further confirms that patients with combined treatment will remain in Stage 1 longer than those with single treatment; however, for patients who had relapse of breast cancer, patients with single treatment will stay alive longer than those with combined treatment.

Another goal of this study was to provide a transition probability matrix at different times so that given a specific time period, the probability that a patient in a given stage will transit to another stage could be conveyed. Tables 5a-8b give 2-year, 4-year, 5-year and 10-year transition probability matrixes of patients in RT+Tam and Tam.

Conclusion

Through Markov chain modeling of the three stages of breast cancer patients , it has been shown that combined treatment of tamoxifen and radiation is more effective than single treatment of tamoxifen in preventing the recurrence of breast cancer. However, for patients who had a relapse of breast cancer, single treatment of tamoxifen proves to be more effective than combined treatment with respect to the survival probability. This finding could give significant guidance to doctors with respect to which breast cancer treatment should be given to breast cancer patients in different stages. Transition probabilities between different stages during 2 years, 4 years, 5 years and 10 years are also calculated for predicting purposes. Those transition probabilities could help provide a clearer view of how one stage transits to another stage within a given time period.

Acknowledgements

We wish to thank N.A Ibrahim for supplying us the source of the data that made the subject study possible. We also wish to express appreciation to Dr. James Kepner, Vice President of the American Cancer Society for our useful discussions on the present study.

References

Dynkin, E. B. (2006). *Theory of Markov Processes*. Dover Publications.

Fyles, A. W., & McCready, D. R. (2004). Tamoxifen with or without breast irradiation in women 50 years of age or older with early breast cancer, *New England Journal of Medicine*, *351*, 963-970.

Gentleman, R. C., Lawless J. F., Lindsey J. C., & Yan, P. (1994). Multi-State Markov models for analyzing incomplete disease history data with illustrations for HIV disease. *Statist. Med*., *13*, 805-821.

Ibrahim N. A., et al. (2008). Decision tree for competing risks survival probability in breast cancer study, *International Journal of Biomedical Sciences*, Volume 3 Number 1.

Jackson, C. (2007). *Multi-State modeling with R: The msm package, version 0.7.4.1 October*.

Kalblfleisch, J. D., & Lawless J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, *80*, 863-871.

Kay, R. (1986). A Markov model for analyzing cancer markers and disease states in survival studies. *Biometrics*, *42*, 855-865.

Lu, Y. & Stitt, F. W. (1994). Using Markov processes to describe the prognosis of HIV-1 infection. *Medical Decision Making*, *14*, 266-272.

Markov, A. (1906). Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga, Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete, 2-ya seriya, tom 15, *9 4*, 135-156.

Norris J. R. (1998). *Markov Chains*, Cambridge University Press.

Satten, G. A. & Longini, L. M. (1996). Markov chains with measurement error: estimating the true course of a marker of the progression of human immunodeficiency virus disease. *Applied Statistician*, *45*, 275-309.

Sharples, L. D. (1993). Use of the Gibbs sampler to estimate transition rates between grades of coronary disease following cardiac transplantation. *Statistics in Medicine*, *12*, 1155-1169.

Table 1: Transition Intensity Matrix of RT+Tam

	Stage 1	Stage 2	Stage 3
Stage 1	-0.02301	0.01957	0.0034
Stage 2	0	-0.3074	0.3074
Stage 3	0	0	0

Table 2: Transition Intensity Matrix of Tam

	Stage 1	Stage 2	Stage 3
Stage 1	-0.03917	0.03528	0.003889
Stage 2	0	-0.08533	0.08533
Stage 3	0	0	0

Figure 3: Survival Curves of Patients in RT+Tam

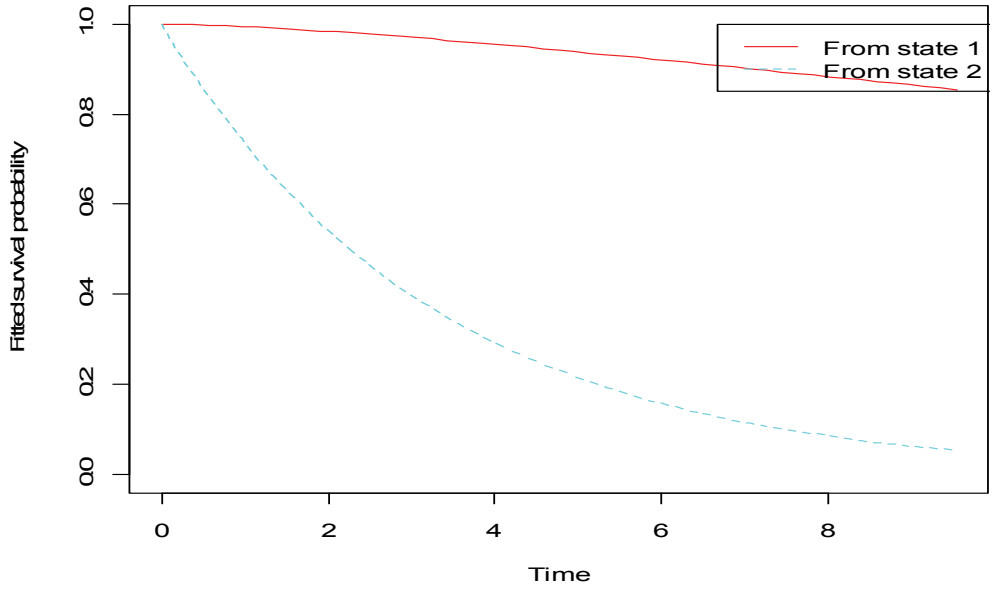


Figure 4: Survival Curves of Patients in Tam

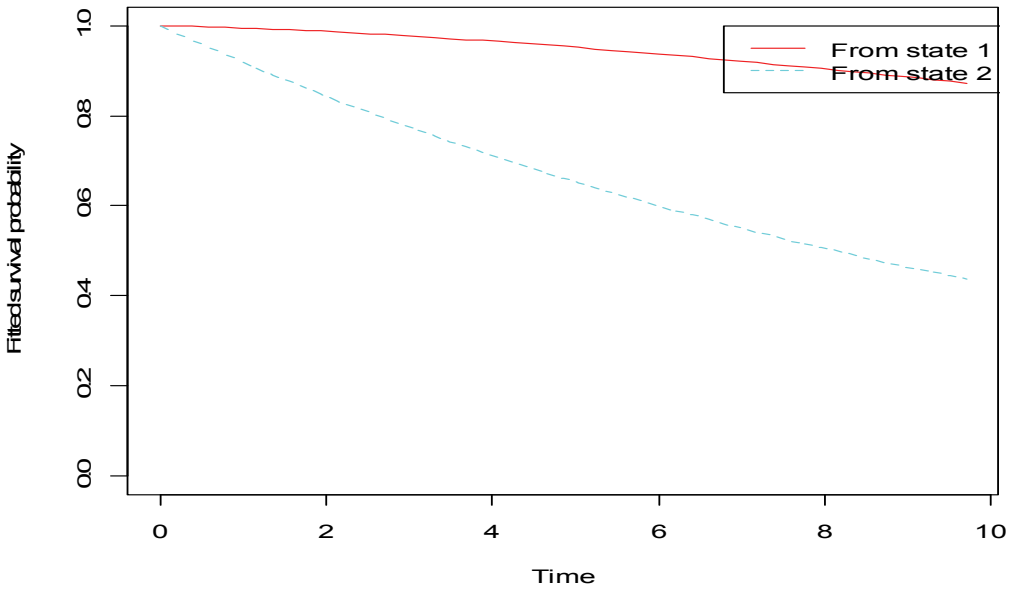


Table 5a: 2-year transition matrix for RT+Tam

	Stage 1	Stage 2	Stage 3
Stage 1	0.9550	0.0285	0.0165
Stage 2	0	0.5408	0.4592
Stage 3	0	0	0

Table 5b: 2-year transition matrix for Tam

	Stage 1	Stage 2	Stage 3
Stage 1	0.9247	0.0623	0.0130
Stage 2	0	0.8431	0.1569
Stage 3	0	0	0

Table 6a: 4-year transition matrix for RT+Tam

	Stage 1	Stage 2	Stage 3
Stage 1	0.9121	0.0426	0.0453
Stage 2	0	0.2925	0.7075
Stage 3	0	0	0

Table 6b: 4-year transition matrix for Tam

	Stage 1	Stage 2	Stage 3
Stage 1	0.8550	0.1102	0.0348
Stage 2	0	0.7108	0.2892
Stage 3	0	0	0

Table 7a: 5-year transition matrix for RT+Tam

	Stage 1	Stage 2	Stage 3
Stage 1	0.8913	0.0466	0.0621
Stage 2	0	0.2151	0.7849
Stage 3	0	0	0

Table 7b: 5-year transition matrix for Tam

	Stage 1	Stage 2	Stage 3
Stage 1	0.8221	0.1295	0.0484
Stage 2	0	0.6527	0.3473
Stage 3	0	0	0

Table 8a: 10-year transition matrix for RT+Tam

	Stage 1	Stage 2	Stage 3
Stage 1	0.7945	0.0515	0.1540
Stage 2	0	0.0463	0.9537
Stage 3	0	0	0

Table 8b: 10-year transition matrix for Tam

	Stage 1	Stage 2	Stage 3
Stage 1	0.6759	0.1910	0.1331
Stage 2	0	0.4260	0.5740
Stage 3	0	0	0