Neural, Parallel, and Scientific Computations 17 (2009) 281 - 290

SURVIVAL AND RELAPSE TIME AMONG DIFFERENT HISTOLOGY TYPES OF BREAST CANCER

CHUNLING CONG[1] AND SARA SAMBANDHAM[2]

[1]Department of Mathematics and Statistics, University of South Florida, FL, 33613
chunlingcong@ifnaworld.org

[2]Department of Medicine, Albany Medical College, Albany, NY, 12208
ssambandham@gmail.com

Abstract.

The objective of the present study is to perform statistical analysis of breast cancer data that involves several different histological types: ductal, lobular, mixed, medullar, mucinous, and others. Survival time and relapse time of two major histology types: ductal and mixed, are analyzed. In addition, we review some of the relevant recent research on the same subject data with respect to parametric analysis, statistical modeling and Markov process.

I. Introduction

Breast cancer is defined by Wikipedia as a cancer that starts in the breast, usually in the inner lining of the milk ducts or lobules. It is the second most common type of cancer worldwide after lung cancer of all cancer incidence, both sexes counted. Therefore, it is of great importance to investigate the variables or factors that contribute to breast cancer, how to predict the survival time or reoccurrence time if it does occur, the probability of cancer transition from one stage to another, and if different histology types of breast cancer affect the survival time and relapse time of breast cancer patients.

The real data is obtained from a study between December 1992 and June 2000 where a total of 641 breast cancer patients at the Princess Margaret Hospital were enrolled and randomized into two different study arms [1]. One arm consists of 320 patients who received combined treatment with radiation and tamoxifen (**RT+Tam**), and the other arm consisting of 321 patients who received single treatment of tamoxifen (**Tam**). The last follow up was conducted in the summer of 2002. Based on the histology type and treatment they received, those 641 breast cancer patients can be divided into the following several subgroups shown in Figure 1 for later analysis. It can be noticed that the majority of the breast cancers are ductal (397) or mixed(174), only a small number are lobular (31), medullar(5), mucinous (16) or others(18).

Received November 11, 2009

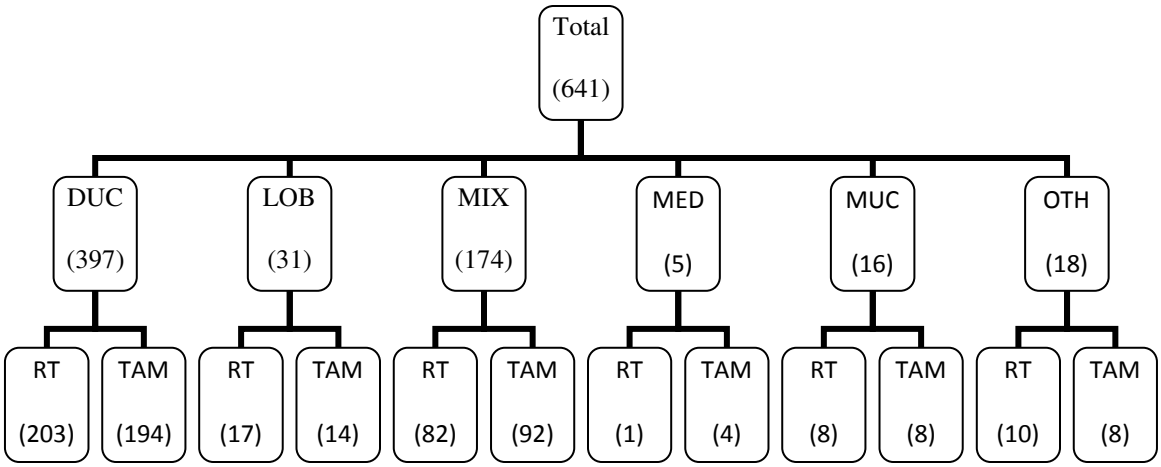


Figure 1. Breast Cancer Patients Grouped by Histological Types and Treatments

Information concerning potential prognostic factors (attributable variables) are **pathsize** (size of tumor in cm); **hist** (Histology: DUC=Ductal, LOB=Lobular, MED= Medullar, MIX=Mixed, MUC=mucinous, OTH=Other); **hrlevel** (Hormone receptor level: NEG=Negative, POS=Positive); **hgb** (Hemoglobin g/l); **nodediss** (Whether axillary node dissection was done: Y=Yes, N=No); **age** (Age of the patient in years). The response variables we are interested in are survival time and relapse time of a given patient.

In the next several sections, related research of breast cancer based on the same dataset is first reviewed, then we proceed to addresses the following questions: is there a difference of survival curves between different histological types; is there a difference of relapse time between different histological types; do patients of different cancer types react differently to treatments with respect to survival time and relapse time?

II. Parametric Analysis & Statistical Modeling

First we would like to review some recent studies that investigated the relapse time of breast cancer patients, of whom one group received combined treatment of tamoxifen and radiation (**RT+Tam**) and other group received single treatment of tamoxifen (**Tam**). For both groups, parametric analysis is conducted to find the best fitted-distribution of the relapse time and the fitted survival curves from the parametric distribution are compared to that of Kaplan-Meier curve with its 95% confidence intervals to show how good the fit is. For **RT+Tam** group, the best fitted probability distribution is found to be the Lognormal distribution with corresponding maximum likelihood estimator (MLE) of the following form: $\hat{\mu}=5.148, \hat{\sigma}=2.47$. For Tam group, Lognormal is also the best fitted distribution with MLE of $\hat{\mu}=3.491$ and $\hat{\sigma}=1.79$. The cumulative distribution function

(CDF) of the fitted distribution is plotted against Kaplan-Meier curve for these two groups in Figure 2 and 3, respectively.

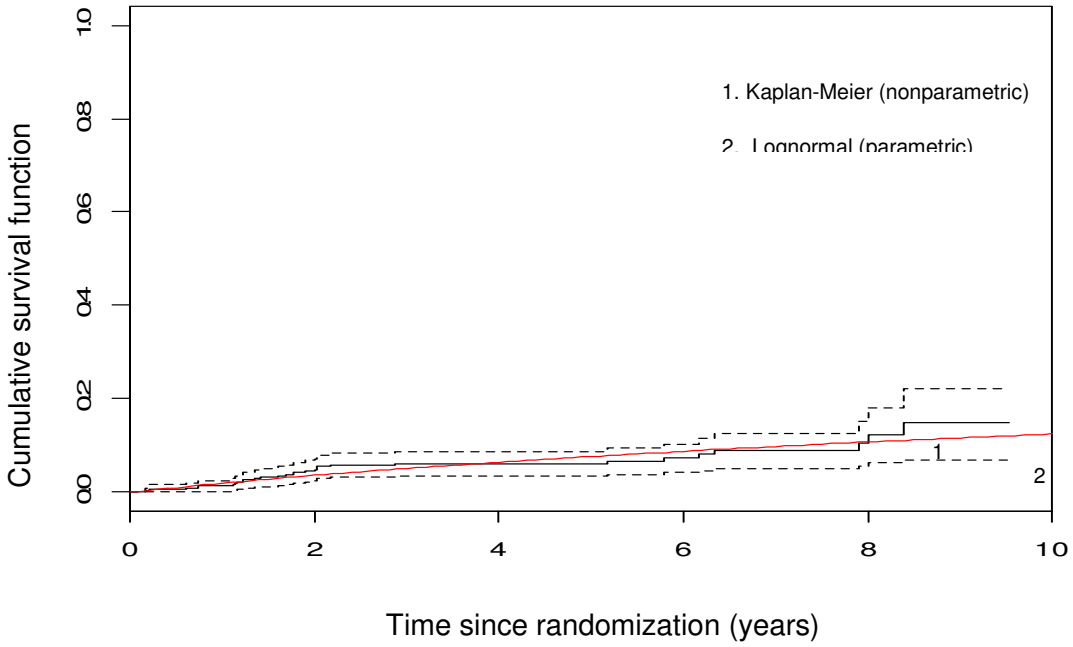


Figure 2. Fitted lognormal CDF curve for RT+Tam

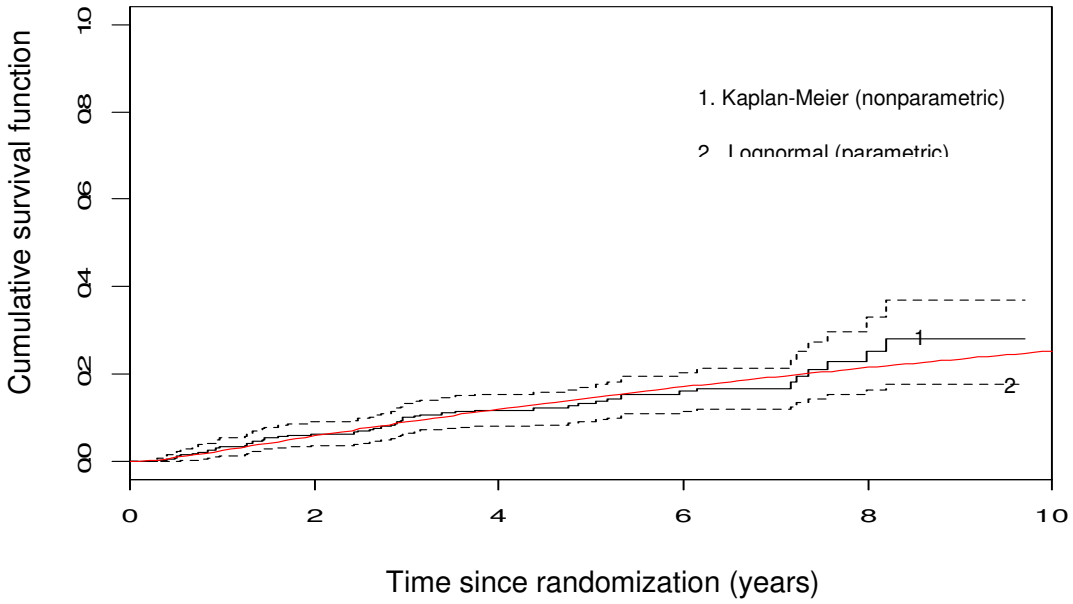


Figure 3. Fitted lognormal survival curve for Tam

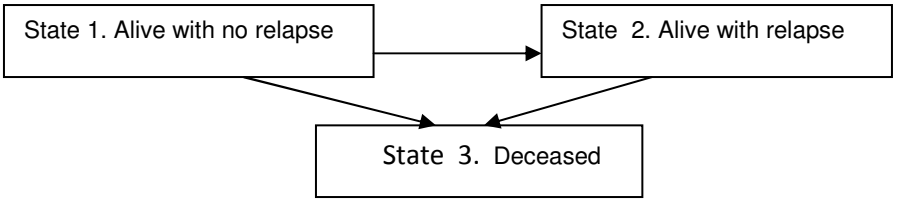
Since relapse time in **RT+Tam** and **Tam** group both follow lognormal probability distribution, the log-likelihood ratio test is applied to test the hypothesis that the mean relapse times of the two groups are equal, in other words, treatment does not affect the mean relapse time. It is shown that there is significant difference for the locations of the two Log-normal distributions between the two treatments groups and it suggests combined treatment is more effective than single treatment in respect to relapse time. More details can be obtained from reference [2].

After knowing the treatment effects to relapse time of breast cancer patients, statistical models are constructed to predict relapse time. For example, given the information of the attributable variables of a given breast cancer patient [3], one would be able to identify how much time it takes before the reoccurrence of breast cancer. Accelerated Failure Time (AFT) [4] model and Cox Proportional Hazard (Cox-PH) [5] model are used to identify the significant attributable variables as well as all possible interactions that contribute to breast cancer relapse time for each treatment group. After running the model including all covariates and interactions between covariates, Arkariki Information Critria (AIC) [6] is used to measure of the goodness of fit of the estimated statistical models.

The results show that for patients who received the combined treatment, **age**, **pathsize**, **nodediss**, **hrlevel**, and the interactions between **age** and **nodediss**, and interaction between **nodediss** and **hrlevel** are significant with respect to relapse time. For patients who received tamoxifen only, **nodediss**, **hrlevel** as single attributable variables, the interactions between **age** and **nodediss**, the interaction between **hgb** and **nodediss** are significant with respect to relapse time in this group.

III. Markov Process

Markov chain is as an efficient way of describing a process in which an individual moves through a series of states in a continuous time has been widely used in health field where the progression of a certain disease are of great importance to both patients and doctors. Thus, this same dataset is analyzed as a Markov Process to investigate the progression of breast cancer patients in three different states who took different treatments [7]. The three different states (alive with no relapse, alive with relapse, and deceased) in the analysis are illustrated in Figure 4 below.



\

Figure 4. Three Stages of Breast Cancer

It was found from Table 1 and 2 through Markov process that patients who received single treatment have a higher transition intensity (0.03528) from State 1 to State 2 than that of combined treatment group which is 0.01957. Thus, those patients in combined treatment group are more likely to have breast cancer reoccurrence. For those patients who died without relapse, there are not much significant differences between the two treatments as illustrated by the intensity form State 1 to State 3 which are 0.0034 and 0.003889 respectively. However, for those who already experienced relapse of breast cancer, patients who received combined treatments are more likely to die than those who received single treatment since the transition intensity from State 2 to State 3 is 0.3074 and 0.08533 for combined treatment and single treatment, respectively. In other words, combined treatment should be chosen over single treatment to avoid reoccurrence, but for those patients who already had breast cancer relapse, it would be advisable to choose single treatment to extend the time from reoccurrence to death.

	State 1	State 2	State3
State 1	-0.02301	0.01957	0.0034
State 2	0	-0.3074	0.3074
Stage 3	0	0	0

Table 1. Transition intensity matrix of RT+Tam

	State 1	State 2	State3
State 1	-0.03917	0.03528	0.003889
State 2	0	-0.08533	0.08533
State 3	0	0	0

Table 2. Transition intensity matrix of Tam

Estimated mean sojourn times in each transient state for patients who received combined treatment and single treatment are obtained to further confirm the conclusion that patients with combined treatment will stay in State 1 longer than those with single treatment, however, for patients who had relapse of breast cancer, patients with single treatment with stay alive longer than those with combined treatment. Furthermore, 2-year, 4-year, 5-year and 10-year transition probability matrixes are constructed to get the transition probability among states at a given time.

IV. Current Work

Despite the usefulness of the previous work done on the dataset, it does not take into consideration of the possible different behavior of different histological breast cancer types. For example, patients with different cancer type would react differently to the same treatments, and also there are potential significant differences among various cancer types with respect to survival time and relapse time. In this study, we divide the dataset into several subgroups based on the histology of the tumors as shown previously, and confine our study to the major two breast cancer types: ductal (DUC) and mixed (MIX) to address the following questions:

1. Is there significant difference for survival time among different histological breast cancer types?
2. Is there significant difference for relapse time among different histological breast cancer types?
3. Do patients with different histological breast cancer types react the same way to treatment with respect to survival time and relapse time?

Survival Time

It is of importance to see if the survival curves of patients in different cancer types are the same. Thus Kaplan-Meier [8] curves are plotted for each of the three major breast cancer types.

Let $S(t)$ be the probability that an individual will not have reoccurrence of breast cancer after time t. For a sample of size n, denote the observed times until death of n sample members as $t_1 \le t_2 \le t_3 \le ... \le t_n$. Then the nonparametric Kaplan-Meier estimator of the survival function is estimated by:

$$\hat{s(t)} = \prod_{t_i \le t} \frac{n_i - d_i}{n_i}$$

where n_i is the number of survivors just prior to time t_i, and d_i is the number of deaths at time t_i.

Kaplan-Meier estimates of the survival curves of relapse time for the two treatment groups are shown in Figure 5.

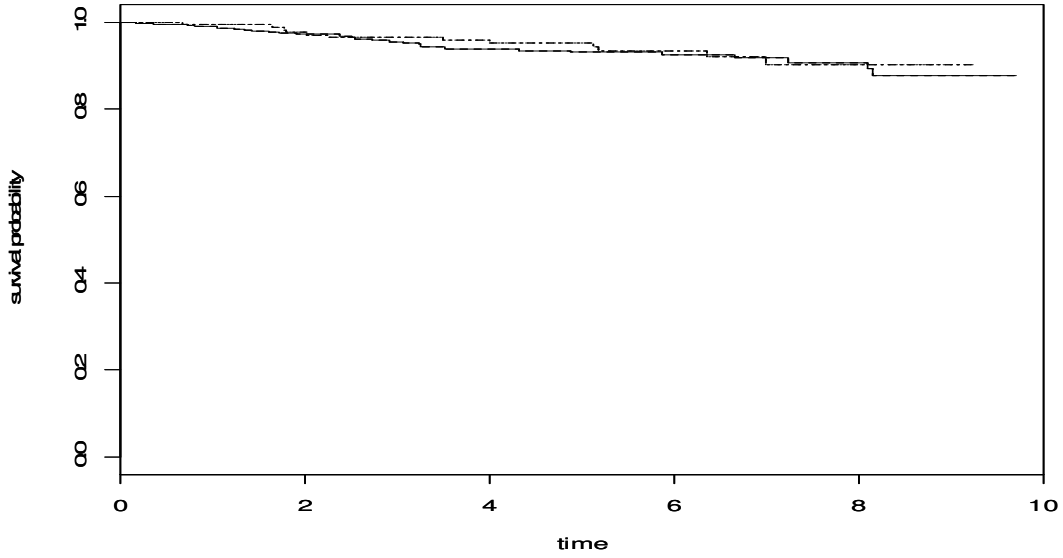


Figure 5. K-M Curves of Survival Times of DUC (solid) and MIX (dotted)

As seen from the graph, the two curves almost overlap showing there is not much difference for survival time of the two breast cancer types. To verify that, Log-rank test [9] is applied and p-value of 0.693 showing that there is no significant difference between survival curves of ductal breast cancer patients and mixed breast cancer patients. This suggests that there is homogeneity of survival time with respect to breast cancer types, so when analysis is conducted on survival time of breast cancer patients, there is no need to separate data into subgroups based on histology type.

Relapse Time

Similar analysis is conducted for relapse time and the Kaplan-Meier survival curves are shown in Figure 6 below.

Furthermore, p-value 0.516 of Log-rank test indicates that there is no significant difference of relapse curve between Ductal and Mixed breast cancer patients.

Treatment

From the previous analysis we find that histology type does not affect the survival and reoccurrence behavior of breast cancer patients. Therefore, we proceed to investigate the treatment effects in different histology types. In another words, we are interested in if combined treatment and single treatment affect survival and relapse time in the same pattern for breast cancer patients with different histological types.

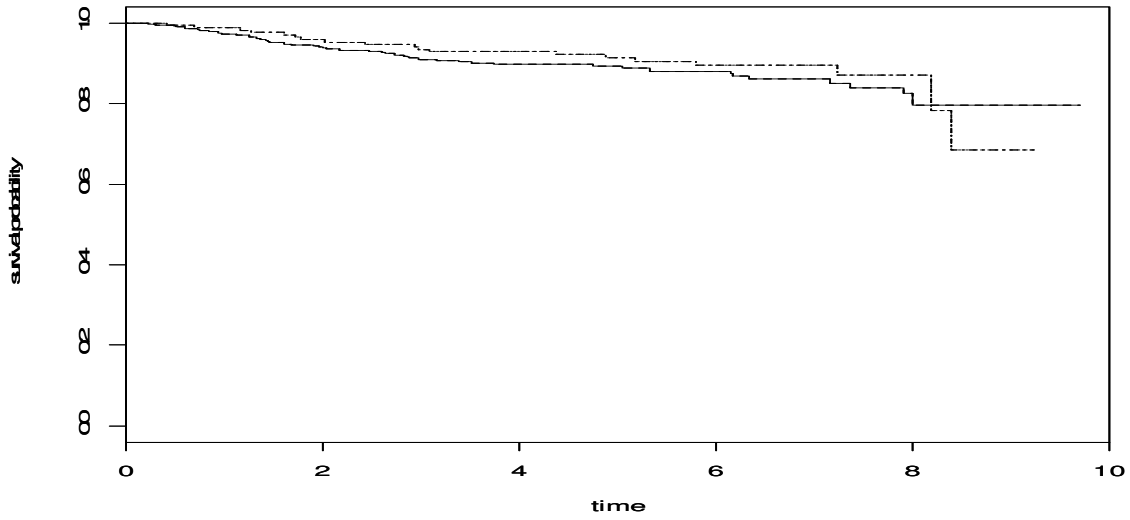


Figure 6. K-M Curves of Relapse Times of DUC (solid) and MIX (dotted)

First, the survival curves of survival time and relapse time of patients of combined treatment group (RT+Tam) and single treatment group (Tam) are compared to see the overall effectiveness of the two treatments. For survival time and relapse time, the Kaplan-Meier curves are shown in Figure 7 and 8 below. And the p-values of the Log-rank test are 0.379 and 0.00192 for survival time and relapse time, respectively. Under significance level of 0.05, it can be concluded that there is no significant difference for survival time between the two treatments. However, combined treatment seems to be more effective than single treatment with respect to relapse time of breast cancer patients.

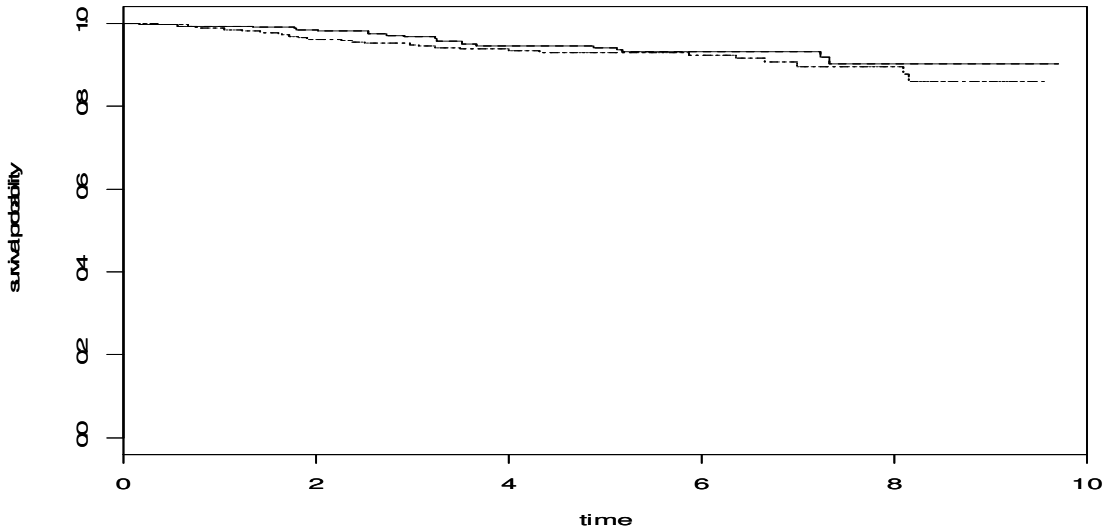


Figure 7. K-M Curves of Survival Times of RT+TAM(solid) and TAM(dotted)

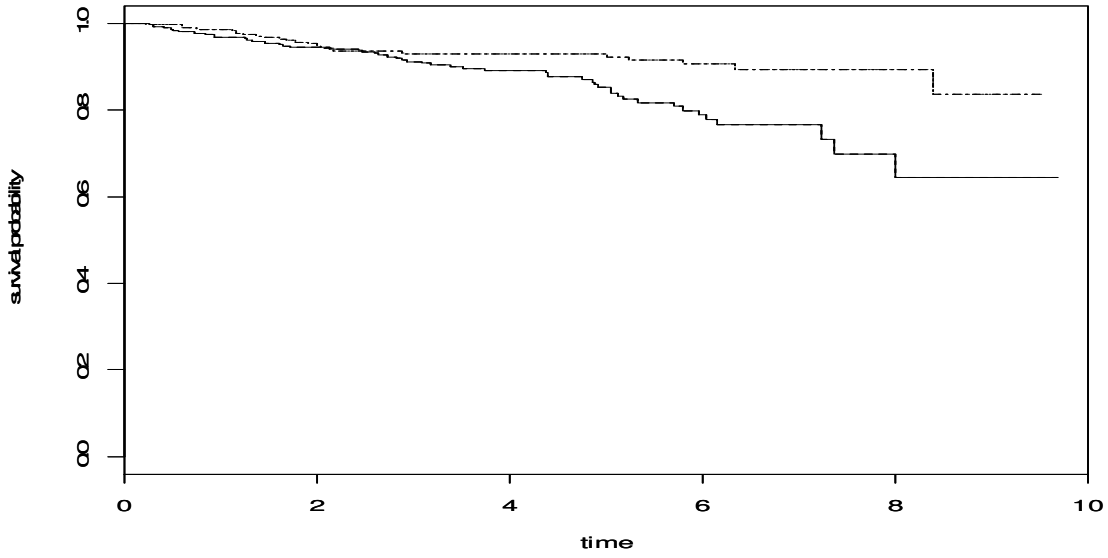


Figure 8. K-M Curves of Relapse Times of RT+TAM(solid) and TAM(dotted)

Same analysis is conducted to the patients groups determined by various histology types. As mentioned above, we confine our analysis to the major two histological types: ductal and mixed because the other histology types do not have enough number of observations for statistical analysis. After running Log-rank test for survival time and relapse time with respect to two different treatments of each histology type, the p-values are obtained and listed in Table 3.

Histology Type	DUC		MIX	
Survival/Relapse	Survival Time	Relapse Time	Survival Time	Relapse Time
P-value	0.217	0.0114	0.708	0.0256

Table 3. Log-rank Test for Survival Time and Relapse Time in DUC and MIX

As can be observed from Table 3, for patients in both DUC and MIX group, survival time of patients who received combined treatment does not significantly differ from those who received single treatment. However, there is significant difference for relapse time between different treatment groups within both DUC and MIX cancer types. This result is consistent with the result obtained from the complete data that consists of all histology types. Thus, breast cancer type does not affect the choice of treatment with respect to survival time and relapse time.

V. Conclusion

Previous research on real breast cancer data is reviewed and the question of homogeneity among different breast cancer histology types is brought into consideration. By dividing the data based on histology type, Kaplan-Meier curve and Log-rank test are performed to test the homogeneity of survival time, relapse time, and treatment effect between the two major histology types: ductal and mixed. Results show there is no significant difference for survival time and relapse time between this two histology types of breast cancer, and treatment effect is the same between the two breast cancer types as well. Thus, there are no significant treatment effects with respect to survival time for DUC, MIX, and the totality data, and combined treatment is more effective than single treatment with respect to relapse time for DUC, MIX and the totality data. These findings provide useful information for statistical analysis and modeling of breast cancer data in the way that all the observations from different histology types can be analyzed as a combined dataset because of the homogeneity among different histology types instead of splitting them into subgroups, and could effectively reduce the time and effort spent on modeling of the subject breast cancer data.

Acknowledgement

The authors would like to acknowledge the useful suggestions and guidance of Dr. Chris P. Tsokos, Distinguished University Professor, University of South Florida, on the subject study.

References

1. A. W. Fyles, D.R. McCready, et al.”Tamoxifen with or without breast irradiation in women 50 years of age or older with early breast cancer”, New England Journal of Medicine 351, pp. 963-970, 2004.

2. C. Cong and C. Tsokos, “Parametric and Nonparametric Analysis of Breast Cancer Treatments”, to appear in “International Journal of Biomedical Sciences”, Volume 5, Number 2, 2010.

3. C. Cong and C. Tsokos, “Statistical Modeling of Breast Cancer Relapse Time with Different Treatments”, submitted for publication

4. Lawless JF. Parametric models in survival analysis. Encyclopedia of Biostatistics. Armitage P. Colton T, Wiley: New York, 3254-64, 2003.

5. Cox, D. R. "Regression Models and Life Tables". Journal of the Royal Statistical Society Series B 34 (2): 187–220, 1972.

6. Akaike H. A new look as the statistical model identification. IEEE Trans Automatic Control, 19: 716-23, 1974.

7. A.A. Markov. "Extension of the limit theorems of probability theory to a sum of variables connected in a chain". reprinted in Appendix B of: R. Howard. Dynamic Probabilistic Systems, volume 1: Markov Chains. John Wiley and Sons, 1971.

8. Kaplan, E.L.; Meier, Paul. “Nonparametric estimation from incomplete observations”. J. Am. Stat. Assoc. 53, 457-481, 1958.

9. Mantel, Nathan . "Evaluation of survival data and two new rank order statistics arising in its consideration.". Cancer Chemotherapy Reports 50 (3): 163–70, 1966.