Journal of the Korean
Data & Information Science Society
2001, Vol. 12, No.2, pp. 71 ~ 82

Tests for Equality of Two Distributions with Life-Table Model

Shin-Soo Kang [1]

Abstract

There are several ways to test the equality of two survival distributions under a variety of situations. Tests for equality of two distributions with life-table model for univariate independent response times are reviewed and introduced. It is developed that the methodology to test it for correlated response times where treatments are applied to different independent sets of cohorts. Data, which can be separated into two independent sets, from an angioplasty study where more than one procedure is performed on some patients are used to illustrate this methodology.

Key Words and Phrases: Interval censoring, Life-table analysis, Product limit estimator, Tests, Wald statistic

1 Introduction

There are several ways to test the equality of two survival distributions, which is often an important issue in the analysis of medical data and in industrial reliability studies. Gehan (1965) adapted the Wilcoxon test to censored data. Cox (1972) considered a two-sample problem with the proportional hazard model and derived an asymptotic two sample test statistic based on a score function. It is similar to test statistics (Mantel and Haenzel, 1959; Mantel,1963) obtained by placing each failed unit in a 2 × 2 contingency table. These statistics are applicable to ranked continuous survival time data assuming no ties in data.

Kalbfleisch and Prentice (1973) developed a discrete model for interval censored data to solve the problem with ties in the Cox model. Thompson (1977) also studied methodology for grouping observations and a logistic model was introduced using

[1] Associate Professor, Department of Information and Statistics, Kwan Dong University, Kangwon-Do, 210-701, Korea

Cox's (1970) binary data methods. The logistic model leads back to the Cox (1972) model as the grouping interval lengths approach zero.

Lininger et al.(1979) considered situations where the experimental units are randomly assigned to one of two treatments within strata which are formed by values of covariates. They compared four test statistics, Gehan, Mantel-Haenzel, continuity corrected Mantel-Haenzel, and Cox through a Monte Carlo study. The continuity corrected Mantel-Haenzel statistic was recommended by Mantel (1966). Note that the Mantel-Haenzel statistic was also called the log rank statistic by Peto and Peto (1972). The Gehan and Mantel-Haenzel statistics tended to have type I error levels closest to the nominal level $\alpha = 0.05$, the continuity corrected Mantel-Haenzel statistic was consistently conservative, with α often less than 0.035. The Gehan statistic was only slightly less powerful than the other statistics. The power of these statistics mainly depended on the total number of failures observed and was otherwise little affected by the degree of censoring or the number of strata. Lininger et al. (1979) suggest that the Cox statistic may be preferable for large data sets with many regression variables.

In small samples, Lininger et al. (1979) concluded that the Cox statistic is not so appropriate as a Mantel-Haenzel statistic because the type I error level for the Cox statistic is not as close to the nominal level $\alpha = 0.05$ as the type I error levels for the Mantel-Haenzel statistics. Farewell and Dahlberg (1983) showed that this result is due to the need to estimate a large number of nuisance parameters in the approach based on the Cox model. Thus, Farewell and Dahlberg (1983) suggest another method that does not require such estimation of nuisance parameters and they show that the modified model perform similarly to a Mantel-Haenzel statistic in small samples. All of these exhibit similar performance in large samples. Little work has been done on comparison of survivor functions using non-parametric estimates involving correlated response times.

Methods for comparing non-parametric estimates of two survivor functions obtained from interval censored data are developed in this paper. In section 2, we review methods for the case where all response times are independent. Two different situations for the correlated response times can be considered: First, there are two independent sets of subjects, each set has independent groups and there are correlated individuals within groups. In this situation, each set of subjects is given a different treatment or each set of subjects is distinguished by a different level of a covariate such as sex or a certain medical condition. This case is discussed in section 3 and an example is given in section 4. The other case is where individuals within groups are randomly assigned to two different treatments. This case is briefly considered in section 5.

2 Univariate Independent Response Times

The simplest way to test the equality of two survivor functions uses the asymptotic normality for the estimates of the survival probabilities to obtain a Wald test as shown in section 3. Other methods presented by Lawless (1982) are reviewed in this section. The methods we consider here are extensions of univariate life-table methods to include a dummy regression variable. We define a dummy regression variable x that has the values 0 or 1 according to whether an individual comes from the first or second population. We now define the following quantities:

N_{ih} = Number of observations "at risk" at time t_{h-1} from the i^{th} population,

D_{ih} = Number of failures in $(t_{h-1}, t_h]$ from the i^{th} population,

C_{ih} = Number of withdrawals in $(t_{h-1}, t_h]$ from the i^{th} population,

P_{ih} = Pr(an individual survives past t_h | the individual is from the i^{th} population),

p_{ih} = $\frac{P_{ih}}{P_{i(h-1)}}$
= Pr(an individual survives past t_h | the individual survives past t_{h-1} and the individual is from the i^{th} population).

For the time being we assume that the censoring events C_{ih} can only occur at the end of the time interval $(t_{h-1}, t_h]$. The observed number of failures in $(t_{h-1}, t_h]$, D_{ih} are binomially distributed random variables with parameters $(N_{ih}, 1 - p_{ih})$. Under these assumptions, the likelihood function is proportional to

$$\prod_{h=1}^{m} \left\{ [1 - p_{1h}]^{D_{1h}} p_{1h}{}^{N_{1h} - D_{1h}} [1 - p_{2h}]^{D_{2h}} p_{2h}{}^{N_{2h} - D_{2h}} \right\}. \tag{1}$$

Let N_h be the risk set at time t_{h-1}, D_h be the set of individuals observed to fail in $(t_{h-1}, t_h]$, C_h be the set of individuals censored in $(t_{h-1}, t_h]$, and P_h be the survival probability beyond $(t_{h-1}, t_h]$. Suppose we have time intervals $(t_{h-1}, t_h]$, for $h = 1, 2, \cdots, m$. Then (1) can be expressed as

$$\prod_{h=1}^{m} \left\{ \prod_{l \in D_h} [1 - p_h(\mathrm{x}_l)] \prod_{l \in N_h - D_h} p_h(\mathrm{x}_l) \right\}, \tag{2}$$

where $p_h(0) = p_{1h}$ and $p_h(1) = p_{2h}$. Brown (1983) also used the binomial distribution of D_{ih} to derive a likelihood function.

Cox (1972) suggested the proportional hazards regression model with conditional hazard function

$$h(t \mid \mathrm{x}) = h_0(t) \exp(\mathrm{x}\beta), \tag{3}$$

and conditional survivor function

$$S(t \mid \mathrm{x}) = [S_0(t)]^{\exp(\mathrm{x}\beta)}, \tag{4}$$

where x is a vector of regression variables, β is a vector of regression coefficients, and the baseline survivor function is $S_0(t) = \exp\left\{-\int_0^t h_0(u)\,du\right\}$. Then (2) can be obtained from (3) by assuming that the lifetime of an individual with regression variable x comes from a proportional hazard model. From (4),

$$\begin{aligned} P_h(\mathrm{x}) &= \Pr(\text{an individual survives past } t_h \mid \mathrm{x}) \\ &= S(t_h \mid \mathrm{x}) \\ &= S_0(t_h)^{\exp(\mathrm{x}\beta)} = P_h^{\exp(\mathrm{x}\beta)}. \end{aligned} \tag{5}$$

Then,

$$\begin{aligned} p_h(\mathrm{x}) &= \frac{P_h(\mathrm{x})}{P_{h-1}(\mathrm{x})} \\ &= p_h^{\exp(\mathrm{x}\beta)}, \end{aligned} \tag{6}$$

where $p_h = p_h(0)$. Then the likelihood function (2) becomes

$$\prod_{h=1}^{m} \left\{ \prod_{l \in D_h} [1 - p_h^{\exp(\mathrm{x}\beta)}] \prod_{l \in N_h - D_h} p_h^{\exp(\mathrm{x}\beta)} \right\}. \tag{7}$$

This approach was introduced by Kalbfleisch and Prentice (1973).

Another model can be applied as following. Let r_h be $\frac{1-p_h}{p_h}$ and

$$p_h(\mathrm{x}) = [1 + r_h \exp(\mathrm{x}\beta)]^{-1}, \tag{8}$$

then

$$\log\left(\frac{1 - p_h(\mathrm{x})}{p_h(\mathrm{x})}\right) = \log r_h + \mathrm{x}\beta. \tag{9}$$

Note that this is a logistic model. Lawless (1982) notes that this model is flexible and convenient, as logistic models are in many other discrete data situations. Prentice and Gloeckler (1978) used a $\log(-\log p_h)$ reparameterization.

A test that survival probabilities are the same for the two populations can be based on a score function. From the logistic model,

$$\begin{aligned} p_{1h} &= p_h(0) = (1 + r_h)^{-1}, \\ p_{2h} &= p_h(1) = (1 + r_h e^{\beta})^{-1}. \end{aligned} \tag{10}$$

Therefore, the null hypothesis $H_0 : p_{1h} = p_{2h}$ is equivalent to $H_0 : \beta = 0$. Under $H_0 : \beta = 0$, the test statistic derived from the score function is

$$Z = \frac{U(0)}{[I(0)]^{0.5}}, \tag{11}$$

where $U(0)$ is the first partial derivative of the loglikelihood function with respect to β, evaluated at $\beta = 0$, and $I(0)$ is the second partial derivative evaluated at $\beta = 0$. Under H_0, Z^2 is approximately distributed as a $\chi^2(1)$ random variable for large sample size.

An alternative method is to replace (2) by a partial likelihood suggested by Cox (1972). The partial likelihood is

$$\prod_{h=1}^{m} \left\{ \frac{\exp(\mathrm{s}_h\beta)}{\left(\sum_{l\in N_h} \exp(\mathrm{x}_l\beta)\right)^{d_h}} \right\}, \tag{12}$$

where d_h is the number of individuals observed to fail in $(t_{h-1}, t_h]$ and $\mathrm{s}_h = \sum_{l\in D_h} \mathrm{x}_l$. Using the score function based on (12), we can test the equality of two distributions. Lawless (1982) said that the two models are nearly equivalent and the tests are virtually identical when intervals are short.

In the previous procedures, we have assumed that all censoring takes place at the ends of the time intervals. If censoring times are uniformly distributed within time intervals, Thompson (1977) suggests the following procedure. Define $G_h = N_h - D_h - C_h$. Then the likelihood (2) is replaced by

$$\prod_{h=1}^{m} \left\{ \prod_{l\in D_h} [1 - p_h(\mathrm{x}_l)] \prod_{l\in G_h} p_h(\mathrm{x}_l) \prod_{l\in C_h} p_h(\mathrm{x}_l)^{0.5} \right\}, \tag{13}$$

and the loglikelihood of (13) is

$$\sum_{h=1}^{m} \left\{ \sum_{l\in D_h} [1 - p_h(\mathrm{x}_l)] + \sum_{l\in G_h} p_h(\mathrm{x}_l) + 0.5 \sum_{l\in C_h} p_h(\mathrm{x}_l) \right\}. \tag{14}$$

This implies that the contribution to the loglikelihood of censored individuals in $(t_{h-1}, t_h]$ is halved. Under $H_0 : p_h(0) = p_h(1)$, the maximization of (14) gives the estimates $\hat{p}_h = 1 - \frac{d_h}{n_h - 0.5c_h}$, which is the AL estimate, where d_h, n_h, and c_h are the number of individuals in D_h, N_h, and C_h, respectively.

3 Correlated Response Times where Treatments are applied to Different Independent Sets of Cohorts

Suppose that there are n independent groups, and some of the groups are randomly assigned to one treatment, while the others are assigned to a second treatment.

Within each group, the individuals may give correlated responses. Instead of exact response times, we have interval censored data. In this situation, it is often required to compare the effects of treatments. For this reason, we sometimes need to test the equality of two survivor functions.

The methods introduced in section 2 can not be applied in this situation because they are based on the independent responses. However, Kang(2000) showed the asymptotic normality of PL and AL estimates of marginal survival probabilities for interval censored bivariate life-times like this situation. Using this property we can compare two survival functions. Let $P1$ be the vector of true survival probabilities under the first treatment and let $P2$ be the corresponding vector of survival probabilities under the second treatment. $\hat{P}1$ and $\hat{P}2$ denote vectors of either the PL or AL estimates from Kang(2000). Covariance matrices of the $\hat{P}i$'s, $V(\hat{P}i)$, i=1,2, are derived using modified Greenwood formula by Kang & Koehler(1997)

Then a test $H_0 : P1 - P2 = 0$ is derived from a Wald statistic as follows. Since $\hat{P}1 - \hat{P}2$ has an asymptotic normal distribution with mean 0 and covariance matrix $\Sigma = V(\hat{P}1) + V(\hat{P}2)$ under H_0, the test statistic

$$Z^2 = (\hat{P}1 - \hat{P}2)' \hat{\Sigma}^{-1} (\hat{P}1 - \hat{P}2), \tag{15}$$

where $\hat{\Sigma}$ is the estimator for Σ computed from the estimates given in Kang & Koehler(1997) for either the PL estimates or AL estimates of $P1$ and $P2$. $\hat{\Sigma}$ is a consistent estimator if the appropriate assumptions about the censoring distribution are satisfied.

Then, when H_0 is true, Z^2 has a χ^2 distribution with m degree of freedom, where m is the rank of $\hat{\Sigma}$.

4 Example for Section 3

The angioplasty data analyzed in Kang & Koehler(1997) can be separated into two independent sets by any one of two variables: gender (male or female) and presence or absence of diabetes mellitus (DM). We will make separate comparisons of survivor functions for each of these two factors.

In Figures 1 and 2, PL estimates of two survival curves are shown for each of the gender and diabetes mellitus factors. Figures 3 and 4 show corresponding curves by AL estimates of survival curves. Although the Figures suggest different survival curves for two levels of factor, tests based on (15) shown in Tables 1 and 2 reveal no significant difference for any of the four factors. This is a result of the high failure rates in the first 9 months and the high levels of censoring. Consequently, there are relatively few individuals at risk after months where the survivor curves appear to be most different, and variances of the estimated survival probabilities are large.

As expected, values of the test statistics in Tables 1 and 2 are smaller when variances and covariances are estimated with modified Greenwood formulas in Kang

& Koehler(1997) than the unmodified formulas. The differences between using modified and unmodified Greenwood estimates are not large in this case because of the large proportions of singletons in this study.

Table 1: χ^2 test for equality of two survivor functions using PL estimates

Test		Gender	DM
Unmodified	χ^2	12.94	10.92
	p-value	(.114)	(.206)
Modified	χ^2	11.39	8.92
	p-value	(.181)	(.349)

Table 2: χ^2 test for equality of two survivor functions using AL estimates

Test		Gender	DM
Unmodified	χ^2	12.96	10.93
	p-value	(.113)	(.206)
Modified	χ^2	11.67	9.06
	p-value	(.166)	(.337)

5 Correlated Response Times where Different Treatments are Applied within Individual Cohorts

Suppose we can distinguish between response times within groups, for example, son or father, male or female litter mates, husband or wife, and so on. In these cases, positive within group correlations can increase the power of tests for comparing survival probabilities.

Fujii (1989) tried to test for the equality of marginal distributions on positively dependent bivariate survival data for a special parametric model. He considered a bivariate survival model where given the value of a common variable, the conditional survival times X and Y independently follow Weibull distributions with hazard functions $\beta_1 t^r$ and $\beta_2 t^r$, respectively, where β_1, β_2, and r are unknown parameters. Then he proposed a test for β_1 equal to β_2.

Clayton (1978) derived the constant odds-ratio model from the constant hazard ratio model for discrete response times given an example of failure times from father-son pair. He suggested that estimating the association parameter θ in the constant

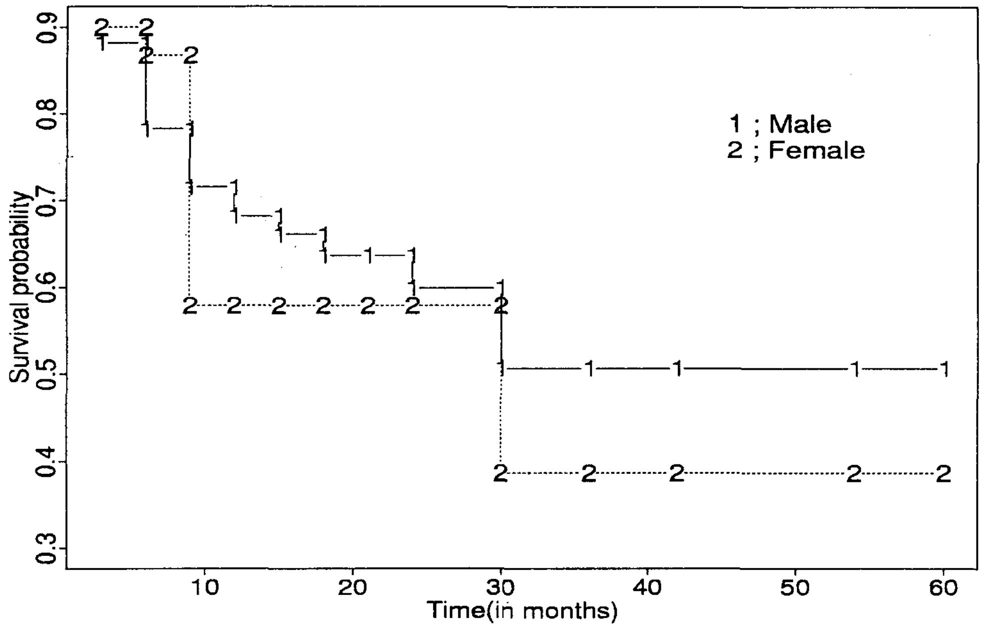

Figure 1: Comparison of PL survival curves for males and females

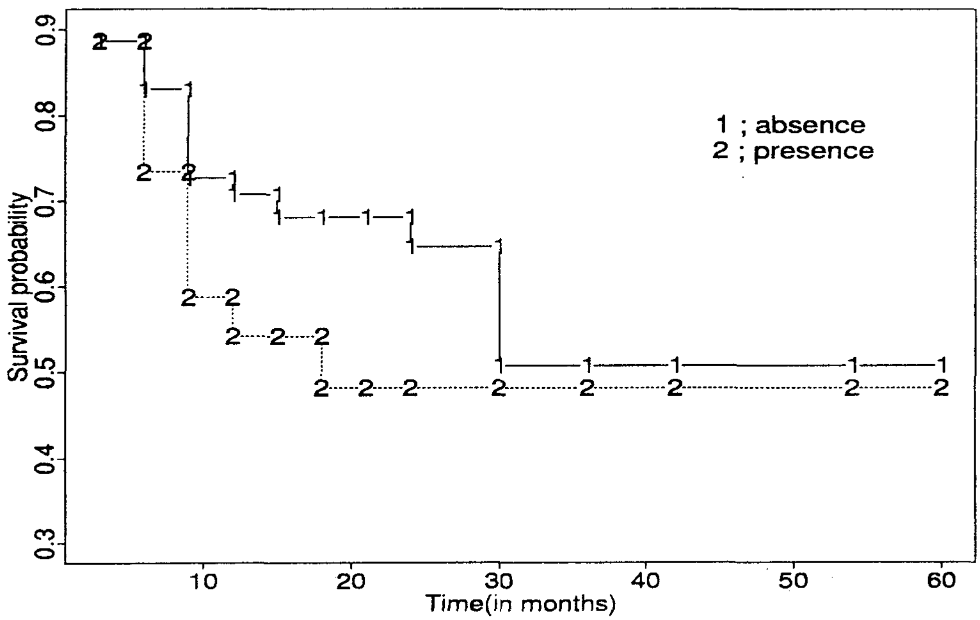

Figure 2: Comparison of PL survival curves for presence and absence of diabetes mellitus

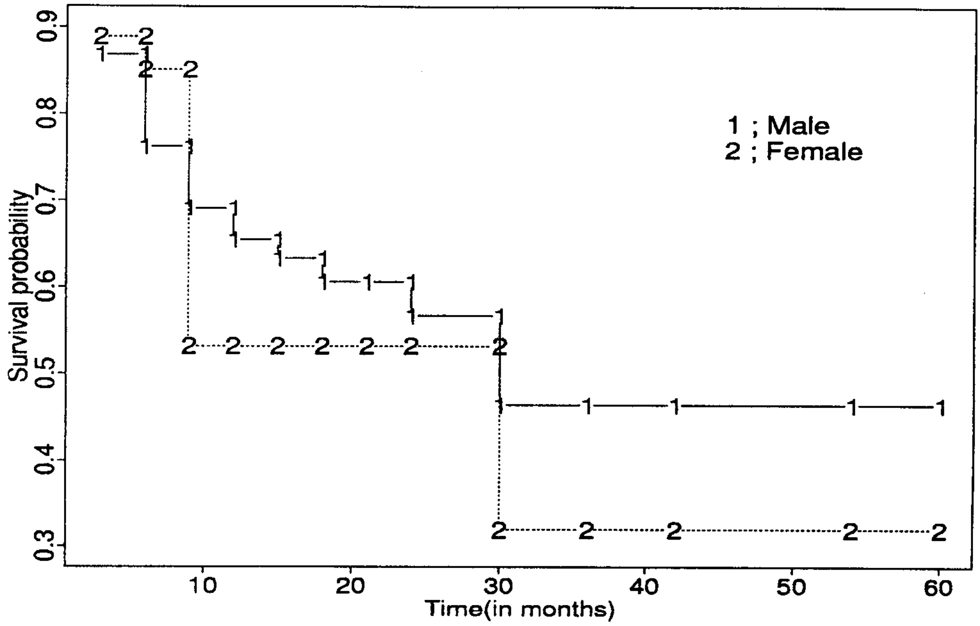

Figure 3: Comparison of AL survival curves for males and females

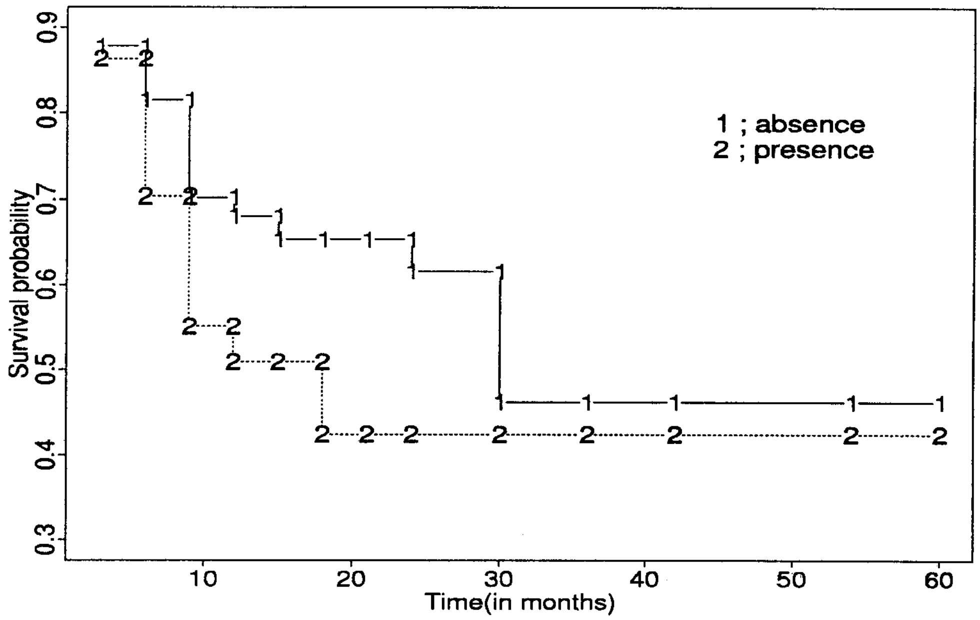

Figure 4: Comparison of AL survival curves for presence and absence of diabetes mellitus

odds-ratio model is asymptotically identical to the Mantel-Haenzel (1959) pooled odds-ratio statistic assuming each pair of trials contributes independently to the information concerning θ. A χ^2 test statistic with 1 degree of freedom for the hypothesis $\theta = 1$, given by Mantel(1963) provides a test of the equality of the survival distributions for fathers and sons.

The development of a Wald test to compare life-table estimators of survival probabilities in this case, is a subject for future research. Let $P1$ is a vector of survival probabilities from the first treatment and let $P2$ be the corresponding vector of survival probabilities from the second treatment. In this case, the estimates $\hat{P}1$ and $\hat{P}2$ are not independent. Consequently, the covariance matrix of $\hat{P}1 - \hat{P}2$ used in (15) is generally not appropriate. Conditional binary responses similar to those introduced in Kang & Koehler(1997) may be used to derive an estimate of the covariance matrix of $\hat{P}1 - \hat{P}2$.

A non-parametric likelihood ratio test for paired data is obtained by testing the fit of the model considered in Kang & Koehler(1998) against a quasi-symmetry model that does not restrict the margins to be homogeneous. For large sample sizes, this would provide a chi-square test with m degrees of freedom when there are $m+1$ time intervals. Evaluation of this test requires development of additional works for maximizing the loglikelihood for the quasi-symmetry model, alternatively, one could compare a model that requires homogeneous margins, but not symmetry, against Campbell's (1981) model.

ACKNOWLEDGEMENTS

This paper was partially supported by Kwan Dong University research fund, 2001. We thank a referee and an associate editor for valuable comments.

References

1. Brown, C. C. (1983). The Statistical comparison of relative survival rates, *Biometrics*, 39, 941-948.

2. Campbell, G. (1981). Non-parametric bivariate estimation with randomly censored data, *Biometrika*, 68, 417-422.

3. Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*, 65, 141-152.

4. Cox, D. R. (1970). *The analysis of Binary Data.* London: Methuen.

5. Cox, D. R. (1972). Regression models and life-tables, *Journal of the Royal Statistical Society, series B*, 34, 187-202.

6. Farewell, V. T. & Dahlberg, S. (1983). On the comparison of procedures for testing the equality of survival curves, *Biometrika*, 70, 707-709.

7. Fujii, Y.(1989). Test for the equality of marginal distributions on positively dependent bivariate survival data, *Communications in Statistics; Simulation*, 18, 633-642.

8. Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly censored samples, *Biometrika*, 52, 203-223.

9. Kalbfleisch, J. D., & Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model, *Biometrika*, 60, 267-279.

10. Kang, S. S. (2000). Asymptotic normality of PL estimator for interval censored bivariate life-times, *The Korean Communications in Statistics*, 7, 245-255.

11. Kang, S. S. & Koehler, K.(1997). Modification of the greenwood formula for correlated response times, *Biometrics*, 53, 885-899.

12. Kang, S. S. & Koehler, K.(1998). Maximum likelihood estimation of bivariate survival probabilities and its existence and uniqueness, *Communications in Statistics; Theory Meth.*, 27(8), 1961-1977.

13. Lawless, J. F. (1982). *Statistical models and methods for lifetime data.* New York: Wiley.

14. Lininger, L., Gail, M., Green, S., & Byar, D. (1979). Comparison of four tests for equality of survival curves in the presence of stratification and censoring, *Biometrika*, 66, 419-428.

15. Mantel, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure, *Journal of the American Statistical Association*, 58, 690-700.

16. Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemotherapy Reports*, 50, 163-170.

17. Mentel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease, *Journal of National Cancer Institute*, 22, 719-748.

18. Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant procedures, *Journal of the Royal Statistical Society, Series A*, 135, 185-206.

19. Prentice, R. L., & Gloeckler, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data, *Biometrics*, 34, 57-67.

20. Thompson, W. A., Jr. (1977). On the treatment of grouped observations in life studies, *Biometrics*, 33, 463-470.