The Korean Journal of Applied Statistics (2009)
22(2), 355–364

Confidence Interval for the Difference or Ratio of Two Median Failure Times from Clustered Survival Data

Seung-Yeoun Lee[1] · Sin-Ho Jung[2]

[1]Department of Applied Mathematics, Sejong University;
[2]Department of Biostatistics and Bioinformatics, Duke University

(Received December 2008; accepted February 2009)

Abstract

A simple method is proposed for constructing nonparametric confidence intervals for the difference or ratio of two median failure times. The method applies when clustered survival data with censoring is randomized either (I) under cluster randomization or (II) subunit randomization. This method is simple to calculate and is based on non-parametric density estimation. The proposed method is illustrated with the otology study data and HL-A antigen study data. Moreover, the simulation results are reported for practical sample sizes.

Keywords: Censoring, cluster randomization, intracluster correlation, quantile, unit randomization.

1. Introduction

Many clinical trials involve the clustered survival data which include times to multiple occurrences of the same type of event. One of important characteristics of this type of data is that times are usually correlated within cluster. There have been many studies for analyzing the clustered survival data that take into account the correlation structure between the various times obtained for the same subject. Xie and Waksman (2003) derived a sample size estimation for the survival times as the primary endpoint to design a clinical trial with the clustered survival data. Recently, issues on developments in design and analysis of cluster randomized trials have been reviewed in a paper celebrating the 25^{th} Anniversary of Statistics in Medicine by Campbell *et al.* (2007). Among many methodological issues in analysing the clustered survival data, we simply focus on comparing two survival curves by estimating the confidence interval of the median survival times in this paper.

In survival analysis, median failure time is often used as a meaningful summary measure to compare the survival experience between different groups. Methods for estimating the confidence interval of the median failure time have been studied extensively for one-sample problem. Brookmeyer and Crowley (1982), Emerson (1982) and Slud *et al.* (1984) developed the procedures by investigating

This work was supported by the faculty research fund of Sejong University in 2007.
[1]Professor, Department of Applied Mathematics, Sejong University, 98 Gunja-dong, Kwangjin-gu, Seoul 143-747, Korea. E-mail: leesy@sejong.ac.kr

Table 1.1. Survival of ventilating tubes for 78 children with otitis media(The paired survival times (in months) are for right and left ears, respectively, and censoring times are marked with + sign)

(i) Control group ($n_1 = 38$)
(3.1, 4.8+), (12.7, 6.0), (3.1, 6.4), (8.5, 12.7), (9.1, 9.1), (0.5, 5.1), (18.1, 15.0), (6.0, 6.0), (6.4, 6.4), (4.4+, 1.3), (12.8, 12.8), (8.8, 8.8), (2.8+, 2.8+), (9.3, 27.1), (6.1, 6.1), (17.9, 20.9), (9.3, 3.1), (2.9, 1.0), (9.1, 9.1), (5.8, 9.3), (2.9, 1.1), (0.8, 8.1), (3.0, 15.8), (9.4, 9.4), (3.1, 3.1), (7.6, 10.1), (5.5, 5.5), (0.7, 0.7), (7.0, 7.0), (11.7, 3.1), (14.3, 3.2)
(ii) Treatment group ($n_2 = 40$)
(11.9, 8.8), (15.4, 9.2), (9.3, 9.3), (15.0, 0.9), (15.0, 11.9), (17.8, 12.2), (5.9, 8.7), (8.9, 12.6), (0.6, 5.7), (6.0, 9.4), (14.6, 9.0), (12.1, 2.9), (3.0, 3.0), (24.9, 8.7), (5.2, 9.0), (24.3, 18.8), (15.2, 12.5), (33.0, 12.1), (13.1, 0.7), (6.1, 17.1+), (9.5, 3.4), (15.1, 17.8), (5.8, 5.8), (0.6, 3.0), (2.8, 1.6), (6.2, 9.0), (8.7, 3.4), (20.9+, 3.4), (9.2, 6.0), (6.4, 14.3+), (8.8, 8.8), (18.5, 13.3), (12.2, 12.2), (12.5+, 8.8), (8.5, 21.7), (1.8, 20.7), (6.2, 9.0), (9.7, 11.1+), (6.0, 6.0), (8.7, 8.7)

Ref. Howie, V. M. and Schwartz, R. H. (1983)

the generalized sign test for the censored data while Efron (1981), Reid (1981) and Hutson (2001) proposed the bootstrap confidence intervals. On the other hand, for two-sample median comparison procedures for censored survival data, Wang and Hettmansperger (1990) developed the method for estimating the confidence interval of the median failure time, in which for the non-shift model, non-parametric density estimation is involved. Su and Wei (1993) developed a minimum dispersion statistic based on Kaplan-Meier estimators and a simple nonparametric confidence interval of the difference or ratio of two median failure times with censoring observations. These methods are based on the assumption that the failure times are independent. However, in many biomedical studies, this independence assumption might not hold. Jung and Su (1993) have generalized the method of Su and Wei (1993) to paired failure times. In this paper, we extend the method of Jung and Su (1993) to the clustered survival data, in which more than two subjects are correlated within a cluster. The form of the variance-covariance matrix of the two Kaplan-Meier estimates is the same except for taking into account of the correlated subjects within a cluster.

First, we consider the cases where a large number of independent clusters are randomized to two arms. Each cluster consists of some (sub)units which are usually correlated each other. For example, in an otology study (Howie and Schwartz, 1983; Teele *et al.*, 1989), 78 children (clusters) suffering from otitis media in the ears (units) received ventilating tubes as a surgical intervention and were randomized to either no treatment ($n_1 = 38$) or a post-surgery treatment ($n_2 = 40$). The paired survival times of tubes were observed from the right and left ears of each child. Children were regularly followed up to check if tubes were functioning or not. One of the study aims was to determine if the treatment extends the life time of tubes or not. Table 1.1 reports the data set taken from Le (1997).

Secondly, in a clinical trial, we may randomize a number of sites from each subject (such as teeth or ears) to different treatments. In this case, the sites (units) share similar characteristics, so that they tend to be dependent, whereas the subjects (clusters) are independent. An example of this type of clustered survival data is taken from a clinical trial conducted to evaluate the influence of HL-A (human lymphocype antigen) on the survival of allogeneic grafts (defined as the time to rejection of graft) in burned patients (Batchelor and Hackett, 1970). The donor and recipient were matched for ABO blood groups and either closely or poorly matched for HL-A antigens. Selected skin-grafts from at least two unrelated typed donners were applied to the patients, so that some of them

Table 1.2. Days of skin graft survival on each burn 16 patients and + indicates censoring

Patient	1	2	3	4	5	6	7	8
Close match	–	24	–	37	19	–	$57^+, 57^+$	93
Poor match	19	–	18,18	29	13	19,19	15	26
Patient	9	10	11	12	13	14	15	16
Close match	16	22	20	18	77,63,29	–	29	60^+
Poor match	11	17	26	21	43	$28^+, 28^+$	15,18	40

Ref. Batchelor, J. R. and Hackett, M. (1970)

involved close HL-A match and the others poor match. The survival times of closely matched and poorly matched skin grafts (units) were observed with both types of graft applied to each patient (cluster). The objective was to discover whether avoiding severe HL-A incompatibility will extend allograft survival time. The survival of allogeneic grafts within each subject tend to be correlated. Table 1.2 reports the data set taken from Batchelor and Hackett (1970).

In this paper, we propose nonparametric confidence intervals for the difference or ratio of two median survival times from clustered survival data under cluster randomization and under subunit randomization, respectively. Our asymptotic results are based on large number of clusters whereas number of subunits within each cluster is bounded. In Section 2, we derive a nonparametric estimation procedure for the confidence intervals of the difference or ratio of two medians in the clustered survival data. In Section 3, we illustrate the proposed procedures using two examples previously introduced. In Section 4, we conduct simulation studies to investigate the performance of the proposed methods for practical sample sizes and a short discussion is given in Section 5.

2. Nonparametric Confidence Intervals for the Difference or Ratio of Two Medians

As described in the previous section, we consider two types of randomization for clustered survival data in estimating the confidence intervals for the difference or ratio of two medians. One is for the clustered survival data under cluster randomization and the other is for the clustered survival data under subunit randomization.

2.1. Confidence interval under cluster randomization

Suppose that n_k clusters are randomized to group $k\,(=1,2)$. Let $n = n_1 + n_2$. For cluster $i\,(=1,\ldots,n_k)$ in group k, let m_{ki} be the number of units, usually called cluster size, and $(T_{kij},\ j = 1,\ldots,m_{ki})$ be their survival times. Units within each cluster have a common marginal survival distribution with survivor function $S_k(t)$ and cumulative hazard function $\Lambda_k(t) = -\log S_k(t)$. We do not specify the marginal and joint distribution functions.

Because of early termination of study or loss to follow-up, the survival times, T_{kij}, may not be completely observed. In conjunction with survival time T_{kij}, let C_{kij} be the censoring time. Then, from units in cluster i of group k, we observe $\{(X_{kij}, \Delta_{kij}),\ j = 1,\ldots,m_{ki}\}$, where $X_{kij} = T_{kij} \wedge C_{kij}$, $\Delta_{kij} = I(T_{kij} \le C_{kij})$ and $a \wedge b = \min(a,b)$. We assume that $(C_{kij},\ j = 1,\ldots,m_{ki})$ are independent of $(T_{kij},\ j = 1,\ldots,m_{ki})$. Note that censoring times may be correlated within each cluster (as in common censoring time case) and may have different distributions in different groups. We assume that the maximum cluster size is bounded, and asymptotic theories in this paper are applied to large number of clusters.

Let $Y_{kij}(t) = I(X_{kij} \geq t)$ and $N_{kij}(t) = \Delta_{kij} I(X_{kij} \leq t)$ be the at-risk and the death processes, respectively, for the j^{th} subunit of cluster i in arm k. Also define $Y_k = \sum_{i=1}^{n_k} \sum_{j=1}^{m_{ki}} Y_{kij}$ and $N_k = \sum_{i=1}^{n_k} \sum_{j=1}^{m_{ki}} N_{kij}$. Let $\hat{S}_k(t)$ be the Kaplan-Meier estimator obtained by assigning equal weight to each subunit. Let $\hat{\theta}_k^p$ be the $100(1-p)\%$ quantile of group k sample, *i.e.*

$$\hat{S}_k\left(\hat{\theta}_k^p\right) = p.$$

With $p = 1/2$, $\hat{\theta}_k^p$ is a sample median. For notational simplicity, we drop the superscript p from $\hat{\theta}_k^p$. Let $y_k(t)$ denote the limit of $n^{-1}Y_k(t)$. Then, by Ying and Wei (1994),

$$\begin{aligned}\sqrt{n}\left(\hat{S}_k(\theta_k) - p\right) &= -p\frac{1}{\sqrt{n}} \sum_{i=1}^{n_k} \sum_{j=1}^{m_{ki}} \int_0^{\theta_k} y_k^{-1}(t) dM_{kij}(t) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n_k} \epsilon_{ki} + o_p(1),\end{aligned}$$

where $M_{kij}(t) = \int_0^t \{dN_{kij}(s) - Y_{kij}(s) d\Lambda_k(s)\}$ and $\epsilon_{ki} = -p \sum_{j=1}^{m_{ki}} \int_0^{\theta_k} y_k^{-1}(t) dM_{kij}(t)$.
Since ϵ_{ki} $(i = 1, \ldots, n_k)$ are independent 0-mean random variables, by the central limit theorem, $\sqrt{n}(\hat{S}_k(\theta_k) - p)$ is approximately normal with mean 0 and variance σ_k^2 that can be consistently estimated by

$$\hat{\sigma}_k^2 = \frac{1}{n} \sum_{i=1}^{n_k} \hat{\epsilon}_{ki}^2,$$

where $\hat{\epsilon}_{ki}$ is obtained from ϵ_{ki} by replacing θ_k, $y_k(t)$ and $M_{kij}(t)$ with $\hat{\theta}_k$, $Y_k(t)/n$ and

$$\hat{M}_{kij}(t) = \int_0^t \{dN_{kij}(s) - Y_{kij}(s) d\hat{\Lambda}_k(s)\}.$$

Here, $\hat{\Lambda}_k(t) = \int_0^t Y_k^{-1}(s) dN_k(s)$ is the Nelson's (Nelson, 1969) estimator of $\Lambda_k(t)$ obtained from the subunits in group k.

Suppose that we are interested in making inferences about $\tau = g(\theta_1, \theta_2)$, for some given function g. For example, τ may be the difference $\theta_2 - \theta_1$ or the ratio θ_2/θ_1. Furthermore, suppose that θ_2 can be expressed as $\theta_2 = h(\tau, \theta_1)$. Consider the quantity

$$W(\tau, \theta_1) = \frac{n\{\hat{S}_1(\theta_1) - p\}^2}{\hat{\sigma}_1^2} + \frac{n\{\hat{S}_2(h(\tau, \theta_1)) - p\}^2}{\hat{\sigma}_2^2}.$$

Then, under $H_0 : \tau = \tau_0$,

$$w(\tau_0) = \inf_{\theta_1} W(\tau_0, \theta_1)$$

asymptotically has a chi-square distribution with 1 degree of freedom. The proof is the same as that given in the Appendix of Su and Wei (1993), except for the form of variance of the two Kaplan-Meier estimates. Hence, we reject $H_0 : \tau = \tau_0$, with significance level α, if $w(\tau_0) > \chi_1^2(\alpha)$, where $\chi_1^2(\alpha)$ is the 100α upper percentile of χ_1^2. Furthermore, $100(1-\alpha)\%$ confidence region for τ is given as

$$\{\tau : w(\tau) < \chi_1^2(\alpha)\}.$$

2.2. Confidence interval under subunit randomization

Suppose that cluster $i\,(=1,\ldots,n)$ has m_i units and m_{ik} of which are assigned to treatment $k\,(=1,2)$, $m_{i1}+m_{i2}=m_i$. Let $T_{ik1},\ldots,T_{ikm_{ik}}$ be survival times for units in treatment k. Since units within a cluster share common characteristics, their survival times $(T_{i11},\ldots,T_{i1m_{i1}},T_{i21},\ldots,T_{i2m_{i2}})$ tend to be positively correlated. Let C_{ikj} be the censoring time. Then, from units in cluster i, we observe $\{(X_{ikj},\Delta_{ikj}),\ j=1,\ldots,m_{ik},k=1,2\}$, where $X_{ikj}=T_{ikj}\wedge C_{ikj}$ and $\Delta_{ikj}=I(T_{ikj}\le C_{ikj})$. We assume that $(C_{ikj},\ j=1,\ldots,m_{ik},k=1,2)$ are independent of $(T_{ikj},\ j=1,\ldots,m_{ik},k=1,2)$. Censoring times may be correlated within each cluster as in common censoring case. Let $Y_{ikj}(t)=I(X_{ikj}\ge t)$ and $N_{ikj}(t)=\Delta_{ikj}I(X_{ikj}\le t)$ be the at-risk and the death processes, respectively, for the jth subunit assigned to arm k from cluster i. Also define $Y_k=\sum_{i=1}^{n}\sum_{j=1}^{m_{ik}}Y_{ikj}$ and $N_k=\sum_{i=1}^{n}\sum_{j=1}^{m_{ik}}N_{ikj}$. Let $\hat{S}_k(t)$ be the Kaplan-Meier estimator obtained by assigning equal weight to each unit.

We assume that within treatment group k, $(T_{ikj},1\le i\le n,1\le j\le m_{ik})$ are marginally identically distributed with survivor function $S_k(t)$ and cumulative hazard function $\Lambda_k(t)=-\log S_k(t)$. By Ying and Wei (1994) ,

$$\begin{aligned}\sqrt{n}(\hat{S}_k(\theta_k)-p)&=-p\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_{ik}}\int_0^{\theta_k}y_k^{-1}(t)dM_{ikj}(t)+o_p(1)\\&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_{ik}+o_p(1),\end{aligned}$$

where $M_{ikj}(t)=\int_0^t\{dN_{ikj}(s)-Y_{ikj}(s)\}d\Lambda_k(s)$, $y_k(t)$ is the limit of $n^{-1}Y_k(t)$ and $\epsilon_{ik}=-p\sum_{j=1}^{m_{ik}}\int_0^t y_k^{-1}(t)dM_{ikj}(t)$. Since $\epsilon_i=(\epsilon_{i1},\epsilon_{i2})^T$ $(i=1,\ldots,n)$ are independent 0-mean random vectors, by the central limit theorem, $\sqrt{n}(\hat{S}_1(\theta_1)-p,\hat{S}_2(\theta_2)-p)$ is approximately normal with mean 0 and covariance matrix V that can be consistently estimated by

$$\hat{V}=\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i\hat{\epsilon}_i^T,$$

where $\hat{\epsilon}_i$ is obtained from ϵ_i by replacing θ_k, $y_k(t)$ and $M_{ikj}(t)$ with $\hat{\theta}_k$, $Y_k(t)/n$ and

$$\hat{M}_{ikj}(t)=\int_0^t\{dN_{ikj}(s)-Y_{ikj}(s)d\hat{\Lambda}_k(s)\},$$

respectively. Here, $\hat{\Lambda}_k(t)=\int_0^t Y_k^{-1}(s)dN_k(s)$ is the Nelson's (Nelson, 1969) estimator of $\Lambda_k(t)$ obtained from the subunits in arm k.

Similarly, let $\tau=g(\theta_1,\theta_2)$ and $\theta_2=h(\tau,\theta_1)$,

$$W(\tau,\theta_1)=n(\hat{S}_1(\theta_1)-p,\hat{S}_2(h(\tau,\theta_1))-p)\hat{V}^{-1}\begin{pmatrix}\hat{S}_1(\theta_1)-p\\ \hat{S}_2(h(\tau,\theta_1)-p\end{pmatrix},$$

$$w(\tau)=\inf_{\theta_1}W(\tau,\theta_1).$$

Then under $H_0:\tau=\tau_0$, $w(\tau_0)=\inf_{\theta_1}W(\tau_0,\theta_1)$ is asymptotically chi-square distributed with 1 degree of freedom. Therefore, $100(1-\alpha)$ % confidence region for τ is given as

$$\{\tau:w(\tau)<\chi_1^2(\alpha)\}.$$

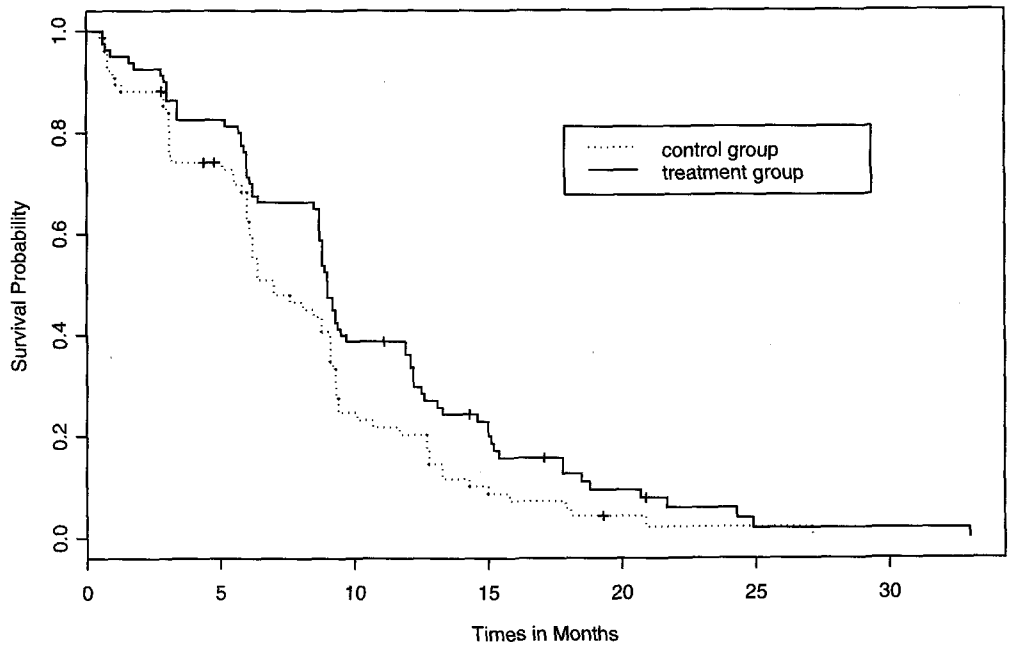

Figure 3.1. Kaplan-Meier curves for control and treatment groups

3. Example

We first apply the proposed method to the otology study (Howie and Schwartz, 1983; Teele *et al.*, 1989) in which 78 children with otitis media were randomized into either no treatment ($n_1 = 38$) or a post-surgery treatment ($n_2 = 40$). The paired survival times of tubes were observed from the right and left ears of each child and shown in Table 1.1. In this case, two arms consist of independent clusters with the paired survival times for each cluster.

As described in Section 2.1, we assume that the paired survival times for right and left ears have a common marginal survival distribution, $S_k(t)$, for the kth treatment group. Let θ_1 and θ_2 be the median failure times for the control and the treatment groups, respectively. The Kaplan-Meier curves of ventilating tubes for two groups are presented in Figure 3.1. The median failure time estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ are 7 and 9 months, respectively. For the parameter $\tau = \theta_1/\theta_2$, the point estimate $\hat{\tau}$ is 0.78, with 95 per cent confidence interval $(0.64, 1.01)$. For the parameter $\tau = \theta_1 - \theta_2$, $\hat{\tau}$ is -2 with 95 per cent confidence interval $(-3.49, 0.09)$. From these two results, the effect of treatment seems to be marginally significant in terms of the median failure time.

The second example is from a clinical trial by Batchelor and Hatckett (1970) as discussed in Section 1. The survival times of closely matched and poorly matched skin grafts (units) were observed with both types of graft applied to each patient, so that the two arms consist of correlated units. Based on the assumption that the marginal survival distributions for the same groups are common, the Kaplan-Meier survival curves for these two groups are presented in Figure 3.2. The survival curve for the closely matched skin grafts seems to be better than that for the poorly matched skin grafts. The median time estimates $\hat{\theta}_1$ for the closely matched group and $\hat{\theta}_2$ for the poorly matched are

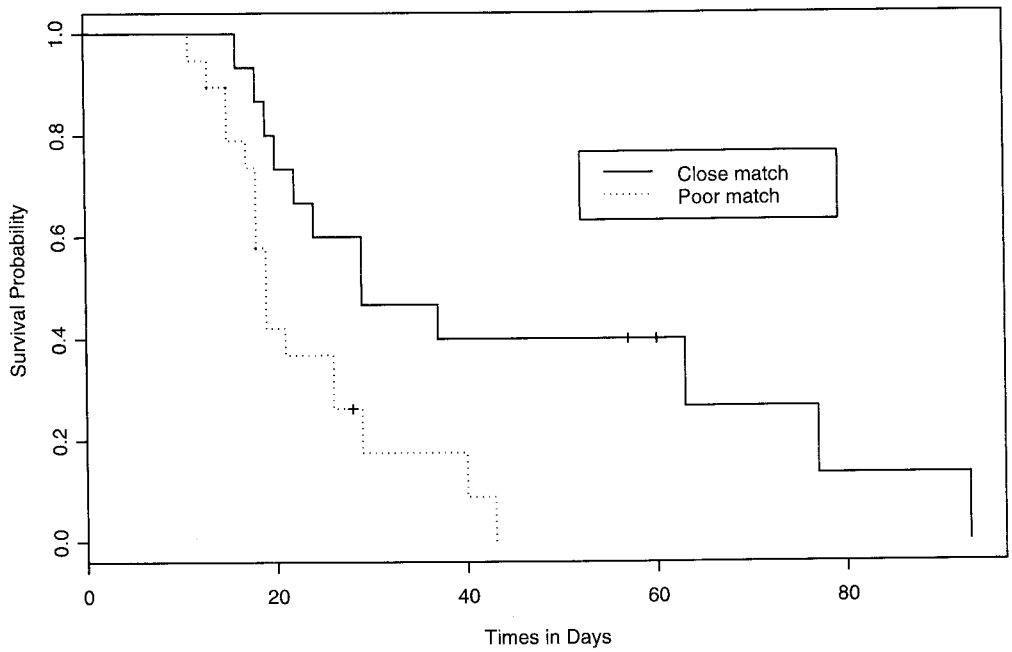

Figure 3.2. Kaplan-Meier curves for close match and poor match groups

29 and 19 days, respectively. For the parameter $\tau = \theta_1/\theta_2$, the point estimate $\hat{\tau}$ is 1.53, with 95 per cent confidence interval (0.94, 3.49). For the parameter $\tau = \theta_1 - \theta_2$, $\hat{\tau}$ is 10 days, with 95 per cent confidence interval $(-1.66, 44.96)$. These results do not provide the significant evidence that avoiding severe HL-A incompatibility will extend allograft survival time. Jung and Su (1993) also analyzed a subset of this data in which only 11 patients are considered since 5 patients has only one matched skin graft available. From their results, the point estimate $\hat{\tau}$ is 1.38, for the parameter $\tau = \theta_1/\theta_2$, with a 95 per cent confidence interval (0.82, 3.75) while for the parameter $\tau = \theta_1 - \theta_2$, $\hat{\tau}$ is 8 days, with 95 per cent confidence interval $(-5.00, 46.00)$. These results lead to the same conclusion with wider confidence intervals by using the partial data.

4. Simulation Results

For investigating the finite sample performance of the proposed methods, simulation studies have been implemented. By choosing $p = 1/2$, we limit our simulations to medians. Table 4.1 reports simulation results under cluster randomization and under dependent subunit randomization. The empirical coverage probabilities of the 95 per cent confidence interval for $\tau = \theta_1/\theta_2$ are presented for a selection of sample sizes, censoring fractions, correlation coefficients and the hazard ratios.

We first generate the clustered survival times under cluster randomization, in which cluster sizes m_{ki} were independently generated from a discrete uniform distribution between 1 and 10. Given cluster size m_{ki}, clustered survival times were generated by modifying Moran's algorithm (Moran, 1967). Let $(U_1, \ldots, U_{m_{ki}})$ and $(V_1, \ldots, V_{m_{ki}})$ be *i.i.d.* normal random vectors with marginally mean 0 and variance 1, and exchangeable structure with correlation coefficient $\sqrt{\rho}$. Then we define

Table 4.1. Empirical coverage of 95 per cent confidence interval for $\tau = \theta_1/\theta_2$(Group 1 hazard rate is set at $\lambda_1 = 1$)

(i) Under cluster randomization

ρ	n_1	n_2	Censoring	$\lambda_2 = 1$	$\lambda_2 = 3/4$	$\lambda_2 = 1/2$
0.0	30	30	30%	.932	.943	.947
			20%	.930	.931	.943
			10%	.924	.936	.941
	50	50	30%	.931	.934	.953
			20%	.927	.939	947
			10%	.929	.933	.942
0.3	30	30	30%	.942	.945	.955
			20%	.931	.941	.943
			10%	.932	.933	.950
	50	50	30%	.937	.943	.948
			20%	.932	.939	.948
			10%	.934	.938	.942
0.6	30	30	30%	.938	.942	.949
			20%	.941	.946	.952
			10%	.938	.942	.950
	50	50	30%	.936	.942	.945
			20%	.937	.945	.947
			10%	.939	.942	.945

(ii) Under subunit randomization

ρ	n	Censoring	$\lambda_2 = 1$	$\lambda_2 = 3/4$	$\lambda_2 = 1/2$
0.0	30	30%	.952	.951	.962
		20%	.948	.955	.960
		10%	.943	.950	.955
	50	30%	.947	.952	.958
		20%	.937	.945	.954
		10%	.936	.944	.947
0.3	30	30%	.957	.963	.966
		20%	.959	.962	.970
		10%	.953	.960	.961
	50	30%	.950	.959	.966
		20%	.953	.957	.965
		10%	.958	.953	960
0.6	30	30%	.973	.978	.973
		20%	.969	.977	.978
		10%	.970	.970	.978
	50	30%	.968	.975	.977
		20%	.967	.968	.980
		10%	.966	.971	.975

$T_{kij} = 0.5(U_j^2 + V_j^2)/\lambda_k, (j = 1, \ldots, m_{ki})$, which have marginally exponential distributions with failure rate λ_k and exchangeable structure with correlation coefficient ρ.

For clustered survival times under unit randomization, we extend Moran's algorithm (Moran, 1967). Let $(U_{11}, \ldots, U_{1m_1}, U_{21}, \ldots, U_{2m_2})$ and $(V_{11}, \ldots, V_{1m_1}, V_{21}, \ldots, V_{2m_2})$ be mutually independent multivariate normal distribution with marginal means of 1, variances of 1 and correlation coefficients, $\mathrm{Corr}(U_{kj_1}, U_{kj_2}) = \mathrm{Corr}(V_{kj_1}, V_{kj_2}) = \sqrt{\rho_1}$, for $j_1 \neq j_2$ and $\mathrm{Corr}(U_{1j_1}, U_{2j_2}) = \mathrm{Corr}(V_{1j_1}, V_{2j_2}) = \sqrt{\rho_2}$. Then for the new random vector $(T_{11}, \ldots, T_{1m_1}, T_{21}, \ldots, T_{2m_2})$ with $T_{kj} = 0.5(U_{kj}^2 + V_{kj}^2)/\lambda_k$,

T_{kj} are marginally exponential distributed with failure rate λ_k and $\text{Corr}(T_{kj}, T_{kj'}) = \rho_1$ for $j \neq j'$ and $\text{Corr}(T_{1j}, T_{2j'}) = \rho_2$. Here ρ_1 and ρ_2 are intracluster correlations between units in the same group and in different groups, respectively. We generate cluster sizes m_i from a discrete uniform distribution between 2 and 5, and allocate these units randomly into two treatments with equal probability.

Censoring variables were generated from *i.i.d.* uniform distribution $U(0, c_0)$ with c_0 chosen for 30% censoring or from $U(c_1, c_0 + c_1)$ with c_1 chosen for 10% or 20% for an exponential survival time with hazard rate 1. Note that c_0 and c_1 may be regarded as accrual and follow-up periods, respectively, in a clinical trial with uniform accrual.

We set ρ and $\rho_1 = \rho_2 = 0$, 0.3 or 0.6, $\lambda_1 = 1, \lambda_2 = 1$, 0.5 or 0.75 and the sample size $n = 30$ or 50. We also set the equal sample sizes $n_1 = n_2 = 30$ or $n_1 = n_2 = 50$ under cluster randomization. For each simulation, 5,000 samples of clustered survival data are generated in each set-up. For each sample, the 95 per cent confidence interval for $\tau = \theta_1/\theta_2$ was constructed. The empirical coverage probability was calculated as the proportion of the 5000 confidence intervals covering the true value $\tau = 1$.

The simulation results given in Table 4.1 show that the empirical coverage probabilities are close to the nominal confidence levels overall. There is no significant difference in coverage probability for various censoring fractions and sample sizes. However, under cluster randomization, coverage probabilities are shown to be anti-conservative when the hazard ratio is one and correlation is rather small, while under subunit randomization, coverage probabilities tend to be conservative when the hazard rate for treatment group is smaller and correlation is large. This trend seems to be caused by the difference of variations between cluster randomization and subunit randomization. Investigating the variance estimates from the simulation results, which are not presented in this paper, the variance estimates under subunit randomization are shown to be larger than those under cluster randomization in all cases.

5. Discussion

We proposed a method for estimating the confidence interval of the median failure times from the clustered survival data. Two examples and simulation results were implemented with visual FORTRAN version 6.1 program. A FORTRAN program for the analysis is available from the authors.

For clustered survival data, two different types of randomization are considered. One is to randomize clusters into two groups while the other is to randomize the subunits from each cluster into two groups. Under cluster randomization, clusters in each treatment are independent while subjects within cluster are correlated. Under subunit randomization, subunits from each cluster share similar characteristics, so that correlated units lie in each group as well as across groups.

As shown in the simulation results, the empirical coverage probabilities are more anti-conservative under the cluster randomization rather than those under the subunit randomization. However, these is no significant difference from the nominal confidence level overall.

References

Batchelor, J. R. and Hackett, M. (1970). HL-A matching in treatment of burned patients with skin allografts, *Lancet*, **2**, 581–583.

Brookmeyer, R. and Crowley, J. (1982). A confidence interval for the median survival time, *Biometrics*, **38**, 29–41.

Campbell, M. J., Donner, A. and Klar, N. (2007). Developments in cluster randomized trials and Statistics in Medicine, *Statistics in Medicine*, **26**, 2–19.

Efron, B. (1981). Censored data and the bootstrap, *Journal of the American Statistical Association*, **76**, 312–319.

Emerson, J. D. (1982). Nonparametric confidence intervals for the median in the presence of right censoring, *Biometrics*, **38**, 17–27.

Howie, V. M. and Schwartz, R. H. (1983). Acute otitis media: One year in general pediatric practice, *American Journal of Diseases in Children*, **137**, 155–158.

Hutson, A. D. (2001). Bootstrap-type confidence intervals for quantiles of the survival distribution, *Statistics in Medicine*, **20**, 1693–1702.

Jung, S. H. and Su, J. Q. (1993). Nonparametric estimation for the difference or ratio of median failure times for paired observations, *Statistics in Medicine*, **14**, 275–281.

Le, C. T. (1997). *Counting Processes and Survival Analysis*, Wiley, New York.

Moran, P. A. P. (1967). Testing for correlation between non-negative variates, *Biometrika*, **54**, 385–394.

Nelson, W. B. (1969). Hazard plotting for incomplete failure data, *Journal of Quality Technology*, **1**, 27–52.

Reid, N. (1981). Estimating the median survival time, *Biometrika*, **68**, 601–608.

Slud, E. V., Byar, D. P. and Green, S. B. (1984). A comparison of reflected versus test-based confidence intervals for the median survival time, based on censored data, *Biometrics*, **40**, 587–600.

Su, J. Q. and Wei, L. J. (1993). Nonparametric estimation for the difference or ratio of median failure times, *Biometrics*, **49**, 603–607.

Teele, D. W., Klein, J. O. and Rosner, B. (1989). Epidemiology of otitis media during the first seven years of life in children in greater Boston: A prospective, cohort study, *Journal of Infectious Diseases*, **160**, 83–94.

Wang, J. L. and Hettmansperger, T. P. (1990). Two-sample inference for median survival times based on one-sample procedures for censored survival data, *Journal of the American Statistical Association*, **85**, 529–536.

Xie, T. and Waksman, J. (2003). Design and sample size estimation in clinical trials with clustered survival times as the primary endpoint, *Statistics in Medicine*, **22**, 2835–2846.

Ying, Z. and Wei, L. J. (1994). The Kaplan-Meier estimate for dependent failure time observations, *Journal of Multivariate Analysis*, **50**, 17–29.