Statistica Sinica 4(1994), 373-387

ASYMPTOTICALLY DISTRIBUTION-FREE MULTIVARIATE RANK TESTS FOR MULTIPLE SAMPLES WITH PARTIALLY INCOMPLETE OBSERVATIONS

Yuko Y. Palesch and John M. Lachin

Medical University of South Carolina and The George Washington University

Abstract: In many clinical trials, it is of interest to compare more than two populations with respect to multiple correlated end-points. In this paper, we present a multivariate rank test for the comparison of R-samples ($R \geq 2$) with respect to multiple time-to-event outcomes as well as to repeated measures. We present a statistic that is a function of a linear combination of stochastic integrals and show that the large sample distribution of a vector of $(R-1)RK$ such stochastic integrals for K ($K > 1$) variates and R groups is asymptotically multivariate normal. We then describe an R-sample T^2-like K-variate omnibus test similar to the Kruskal-Wallis test.

Key words and phrases: Multivariate rank test, counting processes, asymptotic distribution, stochastic integrals, random censorship model, repeated measures analysis, survival analysis.

1. Introduction

Many clinical trials are designed to compare more than two populations with respect to times of multiple events or with respect to a multivariate outcome such as repeated measures over time. For instance, in studies designed to determine the efficacy and/or safety of a new drug using a multivariate outcome variable, the investigators often wish to test for the effects in a placebo group versus perhaps two or more groups receiving varying doses of a drug.

Aalen (1978) showed that the now classical family of weighted Mantel-Haenszel (1959) distribution-free statistics for survival analysis could be obtained using a counting process formulation. Aalen's work was extended into $R > 2$ populations by Andersen, Borgan, Gill, and Keiding (1982) and was used by various authors as the basis for nonparametric R-sample tests (Aalen and Johansen (1978), Gill (1980), Jacobsen (1982, 1984), Harrington and Fleming (1982), Andersen and Borgan (1985), and Jones and Crowley (1989)).

Wei and Lachin (1984) expanded the two-sample univariate Aalen statistic to the multivariate case. Their test can be applied to K-variate time-to-event

data where at most one of the K event types can represent an absorbing state. The Wei-Lachin test allows for general non-informative random censoring where the amount and pattern of censoring may differ between groups provided that censoring occurs at random, i.e. independently of the event times. Wei and Lachin also noted that in the general K-variate case, the model for observations missing at random (completely) (c.f. Little and Rubin (1987)) is a special case of the random censorship model. They then showed that the multivariate rank test can be applied to any multivariate data structure with randomly missing and/or censored observations and is not limited to time-to-event data. See also Davis (1991) and Lachin (1992) for general discussions of the Wei-Lachin and related methods for the analysis of K-variate observations with randomly missing data.

In this paper, we present a multivariate rank test for the R-sample ($R \geq 2$) case. The test is a multivariate extension of the R-sample univariate test of Andersen, Borgan, Gill, and Keiding (1982) using the approach adopted by Wei and Lachin (1984) in their development of a two-sample multivariate test. In Section 2, a brief review of the R-sample univariate test based on the counting process model is given. In Section 3, we develop the R-sample multivariate rank statistic which can be expressed as a linear combination of a larger set of stochastic integrals, and in Section 4, we present a theorem on its asymptotic null distribution and a test of hypotheses based on the multivariate rank statistic. These methods are illustrated in Section 5 with application to data from the National Cooperative Gallstone Study (NCGS) (Schoenfield, et al. (1981)).

2. The R-Sample Univariate Rank Test

We first describe the R-sample ($R \geq 2$) univariate test for a possibly right-censored measure such as an event (survival) time. Let $X_{ij} > 0$ be the time-to-event variable for the jth individual in the ith group. To incorporate information from censored or incomplete observations, let U_{ij} denote a latent variable where $U_{ij} > 0$ under a model where observations may be censored at random, or $U_{ij} \in [0, \infty)$ under a model for missing-at-random. Then, $\tilde{X}_{ij} = \min[X_{ij}, U_{ij}]$, and $\Delta_{ij} = 1$ if X_{ij} was observed and 0 if censored or missing.

Within the ith population, assume that the X_{ij}'s are independently and identically distributed (iid) with survival function $F_i(x)$ and corresponding cumulative hazard function $\Lambda_i(x)$. We wish to test

$$H_0 : F_1(x) = F_2(x) = \cdots = F_R(x) \quad \forall\, x > 0. \tag{2.1}$$

We also assume that the U_{ij}'s are iid as $L_i(u)$ independently of the X_{ij}'s. Note that the censored or missing latent variables may be distributed differently among the populations.

Let the event counting process $N_i(t)$ represent the number of events in the ith group that occur no later than time t:

$$N_i(t) = \#\{j : \tilde{X}_{ij} \leq t, \Delta_{ij} = 1\} = \sum_{j=1}^{n_i} I(\tilde{X}_{ij} \leq t)\Delta_{ij}, \tag{2.2}$$

and let the risk process

$$Y_i(t) = \#\{j : \tilde{X}_{ij} \geq t\} = \sum_{j=1}^{n_i} I(\tilde{X}_{ij} \geq t) \tag{2.3}$$

be the number of individuals at risk in group i at time t. Also, for every i, there exists a left continuous function $y_i(t)$ defined as

$$y_i(t) = [1 - F(t)][1 - L_i(t)] \tag{2.4}$$

such that

$$\sup_{t\in[0,\infty)} \left| \frac{Y_i(t)}{n_i} - y_i(t) \right| \xrightarrow{p} 0 \quad \text{as} \quad n_i \to \infty, \tag{2.5}$$

where $F(t)$ is the assumed common survival distribution under the null hypothesis. Furthermore, let $M_i(t)$ be orthogonal square integrable martingales for group i that can be expressed in terms of counting processes and their compensators:

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)d\Lambda_i(s). \tag{2.6}$$

Finally, let $Q(t)$ be a bounded predictable weighting process that is common to all groups (i) and depends only on the aggregate observable processes, $N.(t) = \Sigma_i N_i(t)$ and $Y.(t) = \Sigma_i Y_i(t)$ (both summations going from $i = 1$ to R), and is equal to zero whenever $Y.(t)$ is zero. Thus, the weight process $Q(t)$ is a function of the observations and vanishes whenever $\min[Y_1(t), \ldots, Y_R(t)] = 0$.

Andersen, Borgan, Gill, and Keiding (1982) proposed a Kruskal-Wallis like test for the null hypothesis (2.1) in terms of the corresponding hazard functions, i.e.

$$H_0 : \Lambda_1(t) = \Lambda_2(t) = \cdots = \Lambda_R(t) \quad \forall \ t > 0. \tag{2.7}$$

Let the elements, Z_i, in the vector $\boldsymbol{Z}$ be

$$Z_i(t) = \int_0^t Q(s)d\Big(\hat{\Lambda}_i(s) - \hat{\Lambda}(s)\Big). \tag{2.8}$$

Thus, $Z_i(t)$ is the integrated weighted difference between the estimated hazard for the ith population and the estimated common hazard under the null hypothesis

(2.7). In practice, these statistics are evaluated at $t = \infty$ or at $t = \max(\tilde{X}_{ij}, i = 1, \ldots, R, j = 1, \ldots, n_i)$.

Algebraically, this expression is equivalent to the weighted Mantel-Haenszel (1959) statistic

$$Z_i(t) = \int_0^t Q(s)\left[dN_i(s) - Y_i(s)\frac{dN_\cdot(s)}{Y_\cdot(s)}\right]. \tag{2.9}$$

By applying the martingale central limit theorem, Andersen, Borgan, Gill, and Keiding (1982) showed that $\boldsymbol{Z}$ is asymptotically normally distributed with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{V}$ that can be consistently estimated by $\hat{\boldsymbol{V}} = \{V_{ir}\}$, where

$$V_{ir}(t) = \int_0^t Q(s)^2 \frac{Y_i(s)}{Y_\cdot(s)}\left[\delta_{ir} - \frac{Y_r(s)}{Y_\cdot(s)}\right] dN_\cdot(s) \tag{2.10}$$

and δ_{ir} is a Kronecker delta or indicator function $\delta_{ir} = I(i = r)$ (see Theorem 3.1 in Andersen et al. (1982)). Then, under the null hypothesis in (2.7), the test statistic $X^2 = \boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}$ is asymptotically distributed as chi-square with $R - 1$ degrees of freedom.

The choice of the weight function $Q(t)$ determines the nature of the test (e.g. modified Wilcoxon or logrank test). The weighted Mantel-Haenszel statistic Z_i includes many families of statistics such as those proposed by Tarone and Ware (1977) and Harrington and Fleming (1982). For example, $Q(t) = 1$ in (2.8) yields the logrank test or the generalized Mantel-Haenszel test. If we set $Q(t) = Y_\cdot(t)$ we have the generalized Gehan's Wilcoxon test proposed by Breslow (1970). If $Q(t) = \hat{F}(t)$, where $\hat{F}(t)$ represents the Kaplan-Meier product limit estimator of the survival function, we obtain the modified Wilcoxon test proposed by Peto and Peto (1972) and Prentice (1978). In general, the choice of a weight process should depend on the nature of the difference in the survival functions $F(s)$ between the groups under the alternative hypothesis of interest.

3. The R-Sample Multivariate Rank Statistics

a. *The Null Hypothesis*: We now describe an R-sample ($R \geq 2$) K-variate ($K > 1$) rank test which can be applied to test the equality of the distributions of K types of events under random censoring or of a K-variate measure under random missingship. The test will be described in the setting of an analysis of multiple events.

An individual j in group i may experience an event of type k at time X_{ijk} and may also experience an event of type ℓ at time $X_{ij\ell}$, where either event may occur before the other and where the occurrence of any one event does not preclude the occurrence of any other events. Thus, a given individual may experience any

number of events between 0 and K. For example, patients with diabetes can be followed to the times of occurrence of any one or all of three microvascular complications – retinopathy, nephropathy, and neuropathy.

It is possible for one of the events to be "absorbing" (e.g. death) if its occurrence implies the occurrence of all of the other events. For example, in a study of times to occurrence of fatal and of non-fatal myocardial infarctions (MI), a fatal MI is an absorbing event in that it precludes observing a subsequent non-fatal MI. However, these data could be analyzed using the times to the occurrence of any MI, fatal or non-fatal, as X_{ij1}, and the time to the occurrence of a fatal MI as X_{ij2}.

Under the random censorship model, the vector $\boldsymbol{X}'_{ij} = [X_{ij1}, \ldots, X_{ijK}]$ represents K separate, possibly censored, event times for the ijth subject. The n_i vectors $\boldsymbol{X}_{ij}$ within each group i are assumed to be iid as $F_i(x_1, \ldots, x_K)$. Similarly, the corresponding censoring vectors $\boldsymbol{U}'_{ij} = [U_{ij1}, \ldots, U_{ijK}]$ are assumed to be iid as $L_i(u_1, \ldots, u_K)$. Under the assumption of random censoring, the $\boldsymbol{U}_{ij}$'s are also assumed to be mutually independent of the underlying event time vectors, $\boldsymbol{X}_{ij}$'s. Often in practice, L_i may be degenerate in R^K, e.g. $U_{ijk} = U_{ij\ell}$ for some ij when the kth and ℓth components in the vector $\boldsymbol{X}_{ij}$ share the same censoring mechanism.

In summary, therefore, the groups may have different patterns of censoring (e.g. $L_{ik} \neq L_{i'k}$ for $i \neq i'$, $1 \leq i < i' \leq R$ and any k), and the censoring mechanisms within a given group may be the same for two different events (e.g. $L_{ik} = L_{i\ell}$ for $k \neq \ell$, $1 \leq k < \ell \leq K$ and any i), but no one event process may serve as a censoring mechanism for any other event process (i.e. X_{ijk} is independent of $U_{ij\ell}$ for $1 \leq k < \ell \leq K$ and all i and j).

A special case of random censorship is random missingship, as in the case of repeated measurements. In this case, if the X_{ijk}'s are positive-valued random variables, then the domain of U_{ijk} now is 0 or ∞. Here, X_{ijk} is missing if $U_{ijk} = 0$, and X_{ijk} is observed if $U_{ijk} = \infty$. The concept readily generalizes to any random variables X_{ijk} which are bounded from below. Therefore, the R-sample multivariate rank test also can be applied to multivariate observations in general, such as repeated measurements.

The hypothesis of interest is

$$H_0 : F_1 = F_2 = \cdots = F_R = F \qquad \forall \ (s_1, s_2, \ldots, s_k)' \in [0, \infty)^K \tag{3.1}$$

where F is a common distribution function for all $i = 1, \ldots, R$. Note that because of the monotonic relationship between the distribution function and the cumulative hazard function, the hypotheses (3.1) for $K = 1$ and (2.7) are equivalent.

b. *The Rank Statistic*: For the kth event type, let the aggregate processes (over the R samples) be defined as $dN_{\cdot k}(t) = \Sigma_i dN_{ik}(t)$ and $Y_{\cdot k}(t) = \Sigma_i Y_{ik}(t)$.

Following Andersen, Borgan, Gill, and Keiding (1982), we can define the weighted Mantel-Haenszel statistic for event type k for each of the R samples as:

$$T_{ik} = \frac{1}{\sqrt{n}} \int_0^\infty Q_k(t) Y_{ik}(t) \left[\frac{dN_{ik}(t)}{Y_{ik}(t)} - \sum_{r=1}^{R} \frac{Y_{rk}(t)}{Y_{\cdot k}(t)} \frac{dN_{rk}(t)}{Y_{rk}(t)} \right], \tag{3.2}$$

where $n = \Sigma_i n_i$, and where $Q_k(t)$ is defined as in Section 2 separately for the kth measure. Note that $\Sigma_i T_{ik} = 0$ for each k and that with no censoring, $Q_k(t) = Y_{\cdot k}(t)$ yields the Kruskal-Wallis test statistic.

4. The Asymptotic Distribution of the R-Sample Multivariate Rank Statistic and the Omnibus Chi-Square Test

In this section, we show that the rank statistics (3.2) can be expressed as linear combinations of a larger set of stochastic integrals. We then present, via a theorem, the large sample distribution of the set of $R(R-1)K$ stochastic integrals from which the asymptotic distribution of the rank statistics is then obtained. The natural test of the null hypothesis (3.1) would be an omnibus large sample chi-square test.

a. *The Multivariate Rank Statistic*: Expanding (3.2) yields

$$T_{ik} = \frac{1}{\sqrt{n}} \left[\sum_{r:r\neq i} \int_0^\infty \frac{Q_k(t) Y_{rk}(t)}{Y_{\cdot k}(t)} dN_{ik}(t) - \sum_{r:r\neq i} \int_0^\infty \frac{Q_k(t) Y_{ik}(t)}{Y_{\cdot k}(t)} dN_{rk}(t) \right] \tag{4.1}$$

Substituting the decomposition (2.6) defined under the null hypothesis, (4.1) can now be expressed as

$$T_{ik} = \sum_{r:r\neq i}^{R} \left[a_{(i)rk} W_{i[r]k} + a_{r(i)k} W_{r[i]k} \right], \tag{4.2}$$

where $a_{(i)rk} = +1$ and $a_{r(i)k} = -1$, $i \neq r$, and where the stochastic integral $W_{i[r]k}$ is defined as

$$W_{i[r]k} = \frac{1}{\sqrt{n}} \int_0^\infty \hat{\mu}_{rk}(t) dM_{ik}(t) \tag{4.3}$$

with

$$\hat{\mu}_{rk}(t) = \frac{Q_k(t) Y_{rk}(t)}{Y_{\cdot k}(r)} \tag{4.4}$$

for all $i \neq r$, $1 \leq i < r \leq R$.

Throughout, we use the notation "$i[r]k$" or "$r[i]k$" where the first subscript is the group indicator for the martingale process $dM_{ik}(t)$. The subscript in the brackets is the indicator for the risk process $Y_{rk}(t)$ involved in the weighting function, $\hat{\mu}_{rk}(t)$. Therefore, $W_{i[r]k}$ is a function of $\hat{\mu}_{rk}(\tilde{X}_{ijk})$ and $W_{r[i]k}$ is a function

of $\hat{\mu}_{ik}(\tilde{X}_{rjk})$. Then, $W_{1[1]k}$ and $W_{1[2]k}$ are not necessarily independent, whereas $W_{1[1]k}$ and $W_{2[1]k}$ can be shown to be uncorrelated since they are functions of variables representing two independent sample groups.

Now, let $\boldsymbol{T}$ denote an $RK \times 1$ vector with elements T_{ik},

$$\boldsymbol{T} = \left[T_{11} \cdots T_{R1} | T_{12} \cdots T_{R2} | \cdots | T_{1K} \cdots T_{RK}\right], \tag{4.5}$$

which, from (4.2), can be expressed as $\boldsymbol{T} = \boldsymbol{A}'\boldsymbol{W}$ in terms of an $[(R-1)RK] \times 1$ vector $\boldsymbol{W}$ and an $[(R-1)RK] \times [RK]$ matrix $\boldsymbol{A}$.

b. *Large Sample Distribution*: We now present a theorem which establishes the asymptotic normality of $\boldsymbol{W}$ and thus of $\boldsymbol{T}$.

Assume the following conditions:

(i) $n_i/n \to \rho_i$ as $n \to \infty$ for $0 < \rho_i < 1$; $i = 1, \ldots, R$, $\Sigma_i \rho_i = 1$;

(ii) There exists for every $k = 1, \ldots, K$, a left-continuous function, $q_k(t)$, with bounded variation in t taking on values in $[0, \infty]$ such that

$$\sup_{t \in [0,\infty)} |Q_k(t) - q_k(t)| \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty;$$

(iii) The support of F_i for all $i = 1, \ldots, R$ is a K-dimensional interval and for any i and j, $\Pr[\Delta_{ijk} = 1 \text{ for all } k = 1, 2, \ldots, K] > 0$.

Theorem. *Under the null hypothesis* (3.1), *and conditions* (i) *through* (iii), *the* $R(R-1)K \times 1$ *vector of stochastic integrals* $W_{i[r]k}$ *in* (4.3) *converges in distribution as* $n \to \infty$ *to a multivariate normal vector, i.e.* $\boldsymbol{W} \xrightarrow{L} MN(0, \Sigma_{\boldsymbol{W}})$, *where the covariance matrix* $\Sigma_{\boldsymbol{W}}$ *is consistently estimated by* $\hat{\Sigma}_{\boldsymbol{W}}$ *with elements*

$$\hat{\mathrm{Cov}}\left(W_{i[r]k}, W_{s[m]\ell}\right) = \begin{cases} \dfrac{n_i}{n} \hat{\sigma}_{i[r]k,[m]\ell}, & \text{if } i = s; \\ 0, & \text{if } i \neq s, \end{cases} \tag{4.6}$$

for all $i \neq r, m$ $(i, r, m = 1, \ldots, R)$ *and* $k, \ell = 1, \ldots, K$, *with*

$$\hat{\sigma}_{i[r]k,[m]\ell} = \frac{1}{n_i} \sum_{j=1}^{n_i} \left[\hat{\mu}_{rk}(\tilde{X}_{ijk})\Delta_{ijk} - \hat{\Psi}_{i[r]k}(\tilde{X}_{ijk})\right]\left[\hat{\mu}_{m\ell}(\tilde{X}_{ij\ell})\Delta_{ij\ell} - \hat{\Psi}_{i[m]\ell}(\tilde{X}_{ij\ell})\right] \tag{4.7}$$

and

$$\hat{\Psi}_{i[r]k}(t) = \sum_{j=1}^{n_i} \frac{\hat{\mu}_{rk}(\tilde{X}_{ijk})\Delta_{ijk} I(\tilde{X}_{ijk} \leq t)}{Y_{ik}(\tilde{X}_{ijk})}. \tag{4.8}$$

The proof of the theorem is presented in the Appendix where we follow the approach taken by Wei and Lachin (1984) who used traditional asymptotic theory.

The multivariate martingale central limit theorem could not be applied due to the multivariate nature of the problem. For each population, the kth counting process, N_{ik}, generates its own σ-algebra, $\mathcal{F}_{ik}$, which is a subset of the probability space, Ω_k. Then, by the Doob-Myer decomposition, for each N_{ik}, we have a martingale M_{ik} expressed in terms of the counting process N_{ik} and its compensator Λ_{ik} with respect to the filtration $\mathcal{F}_{ik}$. In the multivariate setting, it is possible to construct a σ-algebra $\mathcal{F}_i$ that is generated by combining all K σ-algebras $\mathcal{F}_{ik}$. However, since the compensator for N_{ik} is specific to the σ-algebra $\mathcal{F}_{ik}$ generated by the counting process N_{ik}, we have not been able to show that each M_{ik} behaves as a martingale with respect to the all encompassing σ-algebra $\mathcal{F}_i$.

Corollary. *Under the conditions for the theorem, the vector of rank statistics T in (4.5) converges in distribution as $n \to \infty$ to a (singular) multivariate normal vector, i.e. $\boldsymbol{T} \xrightarrow{L} MN(0, \Sigma_T)$, where the rank of the covariance matrix, Σ_T, is at most $(R-1)K$. The covariance matrix, Σ_T, is consistently estimated by*

$$\hat{\Sigma}_T = \boldsymbol{A}'\hat{\Sigma}_W\boldsymbol{A}. \tag{4.9}$$

The corollary is a consequence of the Theorem since $\boldsymbol{T} = \boldsymbol{A}'\boldsymbol{W}$ is a linear function of $\boldsymbol{W}$; and since $\boldsymbol{A}$ is non-stochastic, $\hat{\Sigma}_T = \boldsymbol{A}'\hat{\Sigma}_W\boldsymbol{A} \xrightarrow{p} \boldsymbol{A}'\Sigma_W\boldsymbol{A} = \Sigma_T$.

c. *Omnibus Chi-Square Test*: In general, one might wish to test the null hypothesis (3.1) against the omnibus alternative hypothesis, $H_A : F_{ik}(x) \neq F_{i'k}(x)$ for some $i \neq i'$, and some k. From Corollary 1, the T^2-like test statistic of H_0 versus H_A is

$$\boldsymbol{X}^2_T = \boldsymbol{T}'\Sigma_T^-\boldsymbol{T} \tag{4.10}$$

which is asymptotically distributed as chi-square with degrees of freedom equalling the rank of Σ_T. Note that Σ_T is of rank at most $(R-1)K$ and therefore, a generalized inverse, Σ_T^-, is employed in (4.10).

d. *Remarks*: The elements of the covariance matrix are estimated under the general alternative, but it is easily shown that $\hat{\sigma}_{i[r]k,[m]\ell} \xrightarrow{p} \sigma_{i[r]k,[m]\ell}$ under H_0 in (3.1). If it were desired to estimate the covariance matrix under H_0, rather than using the sample-specific functions $\hat{\Psi}_{i[r]k}$ in (4.8), one would employ

$$\hat{\Psi}_{\cdot[r]k}(t) = \sum_{i=1}^{R}\sum_{j=1}^{n_1} \frac{\hat{\mu}_{rk}(\tilde{X}_{ijk})\Delta_{ijk}I(\tilde{X}_{ijk} \leq t)}{Y_{\cdot k}(\tilde{X}_{ijk})}. \tag{4.11}$$

The theorem has many potential applications. In Palesch (1990) and Palesch and Lachin (1993), the theorem is used to describe the large sample distribution of the family of general relative risk or hazard ratio estimators proposed by Andersen (1983).

Finally, we note that the above tests are based on large sample distribution theory. However, Wei and Lachin (1983) described an extensive small sample simulation of the size and power of the Wei and Lachin (1984) omnibus test for two groups (4.10) and the corresponding univariate tests. They concluded that a sample size of 50 per group is adequate for the tests to be of proper size even with moderate censoring in one or more groups.

5. An Application: National Cooperative Gallstone Study

We illustrate the R-sample multivariate rank tests using data from the National Cooperative Gallstone Study (NCGS) (Schoenfield, et al. (1981)). The NCGS was a three-group ($R = 3$) clinical trial of chenodiol for the dissolution of gallstones in which 916 patients were randomly assigned to receive either placebo, low dose of chenodiol (350 mg/d), or high dose (750 mg/d) for up to two years of follow-up. Two of the suspected side effects which were evaluated were aggravation of biliary symptoms and increase in serum cholesterol levels.

For our purpose, we consider a subgroup of 163 patients (high dose group, $n_1 = 65$; low dose group $n_2 = 50$; and placebo group, $n_3 = 48$) with floating gallstones who were expected to have a higher incidence of efficacy, i.e. dissolution of gallstones. Data from the high dose and placebo groups are presented in Wei and Lachin (1984). The biliary symptom data and the cholesterol data from the low dose group are given in Tables 1 and 2, respectively. (Note that Table 1 of Wei and Lachin (1984) contains typographical errors. The correct data is available from the authors.)

a. *Gallbladder Disease Progression* (*Multistate Survival Data*): These data consist of event-times, X_{ijk}, for each of the two types of event: biliary pain ($k = 1$) and cholecystectomy ($k = 2$). Biliary pain is an acute disabling gallbladder attack which may have resulted in the need to surgically remove the gallbladder (cholecystectomy). Thus, cholecystectomy is an absorbing event; however, by definition, it also implies the occurrence of biliary pain so that $X_{ij1} \leq X_{ij2}$. Censoring could occur if patients prematurely withdrew from the study, or administratively after two years of follow-up. The '+' in the data set designates censoring.

The omnibus multivariate rank test chi-square values with 4 degrees of freedom (df) using the Gehan, Peto-Peto-Prentice and logrank statistics are 9.18, 9.81 and 5.27 with p values of 0.06, 0.04 and 0.26, respectively. The larger test values with the two Wilcoxon tests may be due to the greater incidence of biliary pain reported in the early part of the study for the placebo and low dose groups.

The table below presents the three-sample univariate rank test chi-square (2 df) and p values (in parentheses) for the biliary symptom data:

Event type	Gehan	Peto-Peto	Logrank
Biliary pain	9.15 (0.01)	9.78 (0.01)	5.25 (0.07)
Cholecystectomy	2.84 (0.24)	3.08 (0.21)	1.96 (0.37)

Again, due to the higher incidence of biliary pain at the early stages of the study, the Wilcoxon test values are larger than the logrank test. However, with cholecystectomy, no such trend is evident.

b. *Serum Cholesterol Levels* (*Repeated Measures Data*): Chenodiol inhibits biliary secretion of cholesterol and hepatic conversion of cholesterol to bile acids which suggested that chenodiol might increase the serum cholesterol levels, thereby predisposing the patients to atherosclerosis. Thus, the serum cholesterol levels were routinely monitored. We consider repeated serum cholesterol measurements obtained at 6 months, 12 months, 20 months and 24 months of treatment where some observations are missing at random.

The omnibus multivariate rank test using the Gehan, Peto-Peto-Prentice and logrank statistics yield chi-square values (8 df) of 12.66, 11.27 and 9.72 and p values of 0.12, 0.19 and 0.29, respectively. Therefore, the multivariate rank test with any of the three weight processes results in failure at $\alpha = 0.05$ to reject the null hypothesis.

The table below presents the chi-square (2 df) and p values for the three-sample univariate rank tests for each visit:

Clinic visit	Gehan	Peto-Peto-Prentice	Logrank
6 months	4.14 (0.13)	3.99 (0.14)	2.25 (0.33)
12 months	6.05 (0.05)	5.41 (0.07)	5.26 (0.07)
20 months	3.67 (0.16)	2.25 (0.33)	3.72 (0.16)
24 months	0.21 (0.90)	0.02 (0.99)	2.24 (0.33)

Note the striking differences in the chi-square and corresponding p values of the two Wilcoxon tests versus the logrank test at 24 months.

Appendix: Proof of the Theorem

The proof of the theorem is outlined in the two propositions below and is based in part on the Appendix of Wei and Lachin (1984) and Palesch (1990).

Using ρ_i and $q_k(t)$ as defined in conditions (i) and (ii) of the theorem, and defining $\tilde{y}_{\cdot k}(t) = \Sigma_i \rho_i y_{ik}(t)$, then from (4.4),

$$\mu_{ik}(t) = \frac{q_k(t)\rho_i y_{ik}(t)}{\tilde{y}_{\cdot k}(t)}, \tag{A.1}$$

where

$$\sup_{t\in[0,\infty)} |\hat{\mu}_{ik}(t) - \mu_{ik}(t)| \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty \tag{A.2}$$

for any $i = 1, \ldots, R$, $k = 1, \ldots, K$. Then, define the random variable

$$V_{i[r]jk} = \mu_{rk}(\tilde{X}_{ijk})\Delta_{ijk} - \int_0^{\tilde{X}_{ijk}} \mu_{rk}(t)d\Lambda_k(t) \tag{A.3}$$

for $i \neq r$ $(i, r = 1, \ldots, R)$, $j = 1, \ldots, n_i$, and $k = 1, \ldots, K$.

Proposition 1. Asymptotic Distribution and Variance of $\boldsymbol{W}^*$

Let $\boldsymbol{W}^*$ be an $[(R-1)RK \times 1]$ vector of elements

$$W^*_{i[r]k} = n^{-1/2} \sum_{j=1}^{n_i} V_{i[r]jk} \tag{A.4}$$

$i \neq r$, $i, r = 1, \ldots, R$; $k = 1, \ldots, K$. Then, as $n \to \infty$,

$$\boldsymbol{W}^* \xrightarrow{L} MN(0, \Sigma_{\boldsymbol{W}^*}). \tag{A.5}$$

The covariance matrix $\Sigma_{\boldsymbol{W}^*}$ has elements consistently estimated by

$$\hat{\mathrm{Cov}}\left(W^*_{i[r]k}, W^*_{s[m]\ell}\right) = \begin{cases} \dfrac{n_i}{n}\hat{\sigma}_{i[r]k,[m]\ell}, & \text{for } i = s; \\ 0, & \text{for } i \neq s, \end{cases} \tag{A.6}$$

where $\hat{\sigma}_{i[r]k,[m]\ell}$ is as given in (4.7).

As a linear function of the iid vectors $V_{i[r]j}$, $\boldsymbol{W}^*$ asymptotically is distributed as multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\Sigma_{\boldsymbol{W}^*}$. The detailed proof of the consistency of the estimator $\Sigma_{\boldsymbol{W}^*}$ in (A.6) follows straightforwardly using the approach of Wei and Lachin (1984), and is provided in Palesch (1990).

Proposition 2. Asymptotic Equivalence of $\boldsymbol{W}$ and $\boldsymbol{W}^*$

The vectors $\boldsymbol{W}$ in (4.6) and $\boldsymbol{W}^*$ in (A.5) are asymptotically equivalent, i.e.

$$\left[W_{i[r]k} - W^*_{i[r]k}\right] \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty \tag{A.7}$$

for each $i = 1, \ldots, R$ and $k = 1, \ldots, K$. From (4.3) and (A.4), we have

$$[W_{ik} - W^*_{ik}] = \frac{1}{\sqrt{n}} \int_0^\infty [\hat{\mu}_{rk}(t) - \mu_{rk}(t)]dM_{ik}(t) \tag{A.8}$$

which is the difference of two stochastic integrals of the form $\int HdM$. It can be shown that (A.8) converges in probability to zero as $n \to \infty$ when certain

conditions detailed in Gill (1980) and Wei and Lachin (1984) are met. Then, (A.7) follows, and consequently, the theorem is proved.

Acknowledgement

The authors thank L. J. Wei for many helpful discussions on this topic, and Robert T. Smythe, K. K. Gordon Lan and the referees for helpful comments.

Table 1. Gallbladder disease progression – time (in days) to occurrence of biliary pain ($k = 1$) and to cholecystectomy ($k = 2$) in the low dose ($i = 2$) treatment group in the NCGS (See Wei and Lachin (1984) for the high dose and placebo groups data.)

1	2	1	2	1	2
28	204	230	230	728+	728+
32	264	260	260	728+	728+
60	748+	271	719+	728+	728+
62	727+	281	281	733+	733+
66	738+	350	350	734+	734+
66	262	361	570+	734+	734+
67	727+	370	605+	735+	735+
77	274	497	497	735+	735+
91	728+	529	529	741+	741+
95	810+	609	731+	753+	753+
102	299	632	736+	753+	753+
123	123	714+	714+	759+	759+
183	360+	721+	721+	763+	763+
184	810+	722+	722+	775+	775+
189	729+	722+	722+	812+	812+
191	743+	727+	727+	826+	826+
225	225	728+	728+		

Table 2. Serum cholesterol levels at baseline and changes from baseline at 6, 12, 20, and 24 months of follow-up in the low dose ($i = 2$) group in the NCGS with month of and reason for premature termination (See Wei and Lachin (1984) for the high dose and placebo groups data.)

0 Mo	6 Mo	12 Mo	20 Mo	24 Mo	Termination	
					Mo	Reason
143	−5	.	.	.	8	Dropout
150	12	19	2	11		
152	91	124	.	.	19	Dissolution
156	12	.	46	45		
163	48	41	70	87		
172	32	31	18	11		
172	−14	−1	6	−27		
176	31	.	.	.	9	Cholecystectomy
184	42	41	34	37		
195	51	23	37	−9		
196	22	.	.	.	8	Cholecystectomy
198	23	18	40	39		
199	26	9	15	26		
205	24	14	−10	−7		
212	47	44	31	45		
213	−29	34	31	43		
213	−22	27	31	70		
218	62	61	59	40		
221	−17	−1	−17	−10		
224	53	21	122	49		
226	61	.	.	.	9	Cholecystectomy
227	68	70	54	58		
227	13	60	.	.	12	Dissolution
229	−6	−27	−6	−18		
230	12	53	.	.	17	Dropout
234	−16	3	−54	−64		
235	31	79	18	82		
235	37	5	10	−6		
235	25	26	6	30		
238	−1	−3	.	.	16	Dropout
240	61	.	−3	3		
247	−1	.	.	.	11	Cholecystectomy
249	−40	47	18	3		
254	30	.	.	.	10	Dropout
259	−67	−11	−46	.		
259	−25	.	.	.	7	Cholecystectomy
261	.	.	.	.	7	Dropout
261	12	.	.	.	7	Cholecystectomy
264	−3	55	−7	43		
265	38	27	58	39		
269	10	15	29	31		
271	−6	−22	−28	−14		
273	−37	.	.	.	9	Dropout
284	−12	−7	27	17		
285	30	52	33	63		
288	−34	−31	−10	7		
290	.	.	.	.	4	Cholecystectomy
294	−20	54	−9	.		
300	34	−48	15	−31		
339	−34	−13	0	−39		

References

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701–726.

Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scand. J. Statist.* **5**, 141–150.

Andersen, P. K. (1983). Comparing survival distributions via hazard ratio estimates. *Scand. J. Statist.* **10**, 77–85.

Andersen, P. K. and Borgan, O. (1985). Counting process models for life history data: A review. *Scand. J. Statist.* **12**, 97–158.

Andersen, P. K., Borgan, O., Gill, R. and Keiding, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data. *Internat. Statist. Rev.* **50**, 219–258.

Breslow, N. (1970). A generalized Kruskal-Wallis test for comparing k samples subject to unequal patterns of censorship. *Biometrika* **57**, 579–594.

Davis, C. S. (1991). Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statist. in Medicine* **10**, 1959–1980.

Gill, R. (1980). *Censoring and Stochastic Integrals.* Mathematical Center Tract **124**, Mathematical Center, Amsterdam.

Harrington, D. P. and Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69**, 553–566.

Jacobsen, M. (1982). *Statistical Analysis of Counting Processes. Lecture Notes in Statist.* **12**, Springer-Verlag, New York.

Jacobsen, M. (1984). Maximun likelihood estimation in the multiplicative intensity model: A survey. *Internat. Statist. Rev.* **52**, 193–207.

Jones, M. P. and Crowley, J. (1989). A general class of nonparametric tests for survival analysis. *Biometrics* **45**, 157–170.

Lachin, J. M. (1992). Some large-sample distribution-free estimators and tests for multivariate partially incomplete data from two populations. *Statist. in Medicine* **11**, 1151–1170.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* John Wiley, New York.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst.* **22**, 719–748.

Palesch, Y. Y. (1990). R-sample multivariate rank tests and related estimators of population differences. Ph.D. dissertation, Department of Statistics, The George Washington University.

Palesch, Y. Y. and Lachin, J. M. (1993). Asymptotically distribution-free hazard ratio estimates for multivariate event-time distributions. Manuscript in preparation.

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). *J. Roy. Statist. Soc. Ser.A* **135**, 185–207.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167–179.

Schoenfield, L. J., Lachin, J. M., and the Steering Committee, and the NCGS Group (1981). Chenodiol (Chenadeoxycholic Acid) for dissolution of gallstones: The National Cooperative Gallstone Study. *Ann. Internal Medicine* **95**, 257–282.

Tarone, R. E. and Ware, J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika* **64**, 156–160.

Wei, L. J. and Lachin, J. M. (1983). Two-sample distribution-free tests for censored multivariate observations. Technical Report, The George Washington University, Dept. of Statistics.

Wei, L. J. and Lachin, J. M. (1984). Two-sample asymptotically distribution-free tests for incomplete multivariate observations. *J. Amer. Statist. Assoc.* **79**, 653–661.

Department of Biostatistics, Epidemiology and Systems Science, Medical University of South Carolina, 171 Ashley Avenue, Charleston, SC 29425, U.S.A.
The Biostatistics Center, Department of Statistics/Computer and Information Systems, The George Washington University, Rockville, MD 20852, U.S.A.

(Received April 1992; accepted July 1993)