BIOMETRICS 56, 984–990
December 2000

Nonparametric Rank-Based Methods for Group Sequential Monitoring of Paired Censored Survival Data

Susan Murray

Department of Biostatistics, University of Michigan,
1420 Washington Heights, Ann Arbor, Michigan 48109-2029, U.S.A.
email: skmurray@umich.edu

SUMMARY. This research gives methods for nonparametric sequential monitoring of paired censored survival data in the two-sample problem using paired weighted log-rank statistics with adjustments for dependence in survival and censoring outcomes. The joint asymptotic closed-form distribution of these sequentially monitored statistics has a dependent increments structure. Simulations validating operating characteristics of the proposed methods highlight power and size consequences of ignoring even mildly correlated data. A motivating example is presented via the Early Treatment Diabetic Retinopathy Study.

KEY WORDS: Clinical trial; Correlated times-to-event; Two-sample test; Weighted log rank.

1. Introduction

Paired designs with positively correlated outcomes have historically minimized variability in comparisons; hence, these designs gain power over similarly sized independent group studies. For paired uncensored times-to-event, tests such as Wilcoxon's signed-rank test or the paired t-test are often used. But survival endpoints occurring after long intervals suffer from right censoring. An example of paired censored survival data is found in the Early Treatment Diabetic Retinopathy Study (ETDRS), which enrolled 3711 patients with mild-to-severe nonproliferative or early proliferative diabetic retinopathy in both eyes from April 1980 to July 1985 (ETDRS Research Group 1991a,b). One eye per patient was randomized to early photocoagulation and the other to deferral of photocoagulation until detection of high-risk proliferative retinopathy. The survival endpoint was time to severe visual loss, with loss defined as visual acuity less than 5/200 at two consecutive visits. Because patients were recruited and followed in the ETDRS over 9 years, accumulating patient survival data was periodically monitored to ensure timely detection of treatment effects. Analyses prepared roughly biannually were used by a Data Monitoring Committee (DMC) to determine whether the trial should end early or be continued.

One popular strategy for monitoring patient treatment responses while protecting overall type I error is to use group sequential methods with error spending functions of Lan and DeMets (1983) stylistically modeled as in Pocock (1977) or O'Brien and Fleming (1979). Group sequential methods for weighted log-rank tests with independent groups have been studied extensively (Tsiatis, 1981, 1982; Sellke and Siegmund, 1983; Slud, 1984; Gu and Lai, 1991). However, little group sequential methodology has been developed for paired censored survival data as in the ETDRS. Chang, Hsiung, and Chuang (1997) considered sequential methods for frailty models assuming common pair entry times. A few authors have studied sequential designs for independent groups with multiple correlated censored outcomes. Lin (1991) devised a nonparametric weighted linear rank statistic for monitoring correlated nonidentically distributed censored outcomes across two independent groups, while Muñoz, Bangdiwala, and Sen (1997) proposed parametric models for monitoring correlated pairs of similar censored outcome types across independent groups.

This research presents an adaptation of Gill's (1980) family of weighted log-rank tests, adjusted for correlation within the paired survival random variables and within the paired censoring random variables, and methodology for sequentially monitoring these nonparametric statistics. Related adaptations of rank-based tests in the case of a single analysis have been considered by Dabrowska (1989) and Huang (1999). Section 2 describes the paired weighted log-rank test (PWLR) for a single analysis, the joint sequential distribution of these tests, and related stopping boundaries. Simulations in Section 3 verify the operating characteristics of the recommended sequential monitoring procedure and show consequences of ignoring the pairing in the censored survival outcomes. This section also revisits the ETDRS. A discussion follows in section 4.

2. Joint Sequential Distribution of PWLR Statistics

To understand sequential theory with dependent times to event, an explanation of notation is required. Let $g = 1, 2$ denote treatment group and $i = 1, \ldots, n$ denote either an individual who experiences both treatments, as in the ETDRS, or a matched pair whose members receive opposing treatments, as in a sibling study. These n individuals or n matched pairs are accrued into the trial at times E_{gi} for $i = 1, \ldots, n$ and $g = 1, 2$. In many cases, $E_{1i} = E_{2i}$ is an individual's en-

try time; otherwise, E_{1i} and E_{2i} denote potentially different times. Entry times are assumed to be bounded positive random variables that are identically distributed within group g with distribution function $G_g(e) = P(E_g \le e)$ and $E_{g_1 i_1}$ independent of $E_{g_2 i_2}$ for $i_1 \ne i_2$. Each individual or matched pair denoted by i has two internal correlated survival times T_{gi}, $g = 1, 2$, measured from their entry time. For instance, in the ETDRS, T_{1i} and T_{2i} measure time to vision loss for eyes randomized to deferred and early photocoagulation, respectively, within individual i. For data analyzed at time t, outcomes with $T_{gi} > t - E_{gi}$ have not occurred prior to the analysis time and are censored. Let V_{gi}, $g = 1, 2$, $i = 1, \ldots, n$, be potential censoring times due to random follow-up loss. Aside from potential dependence between E_{1i} and E_{2i}, between T_{1i} and T_{2i}, and between V_{1i} and V_{2i}, it is assumed that E_{gi}, V_{gi}, and T_{gi} are independent for all $g = 1, 2$ and $i = 1, \ldots, n$. Observable random variables for group g at analysis time t are $\{X_{gi}(t), \Delta_{gi}(t)\}$, for all $i = 1, \ldots, n$ such that $E_{gi} \le t$, where $X_{gi}(t) = \min(T_{gi}, V_{gi}, t - E_{gi})$ is the observed time on study at analysis time t and $\Delta_{gi}(t) = I\{T_{gi} \le \min(t - E_{gi}, V_{gi})\}$ is the failure indicator at time t. Indices for calendar time measured from the start of the study and indices for internal patient time measured from entry into the study will frequently be used in combination. The index t will denote calendar time of an analysis and x will index internal patient time.

For each group g and calendar time t, define the number of events occurring no later than internal time x as $N_g(t, x) = \Sigma_{i=1}^n I\{X_{gi}(t) \le x, \Delta_{gi}(t) = 1\}$ for $0 \le x \le t$, the number at risk at x as $Y_g(t, x) = \Sigma_{i=1}^n I\{X_{gi}(t) \ge x\}$, and sample size enrolled $n_g(t) = \Sigma_{i=1}^n I(E_{gi} \le t)$. The number of entered correlated pairs across treatment groups g_1, g_2 for analysis times t_1, t_2 is $n_{g_1 g_2}(t_1, t_2) = \Sigma_{i=1}^n I(E_{g_1 i} \le t_1, E_{g_2 i} \le t_2)$. Often when outcome pairs are attributed to an individual, $n_1(t) = n_2(t) = n_{12}(t, t)$. If at the final analysis time all pairs have been entered, $n_1(t) = n_2(t) = n_{12}(t, t) = n$. However, this method allows individual pair members to remain unentered at the final analysis if the number of entered complete pairs is approaching ∞. Let $J(t, x) = 1$ if $Y_1(t, x)Y_2(t, x) > 0$ and $J(t, x) = 0$ otherwise. Assume a weighting process, $\hat{w}(t, x)$, such that

$$\sup_{x \in [0,t)} |\hat{w}(t, x) - w(t, x)| \xrightarrow{P} 0$$

for constant $w(t, x)$ and that vanishes for $J(t, x) = 0$. Let $n^*(t) = n_1(t)n_2(t)/\{n_1(t) + n_2(t)\}$. At each analysis time t, consider the asymptotically normal family of test statistics

$$\mathcal{T}(t) = \{n^*(t)\}^{\frac{1}{2}} \int_0^\infty \hat{w}(t, u) \Big[\{Y_1(t, u)\}^{-1} dN_1(t, u) - \{Y_2(t, u)\}^{-1} dN_2(t, u) \Big],$$

extended to handle correlation in the paired censored survival times. For a paired log-rank test,

$$\hat{w}(t, x) = \frac{J(t, x)Y_1(t, x)Y_2(t, x)\{n_1(t) + n_2(t)\}}{n_1(t)n_2(t)\{Y_1(t, x) + Y_2(t, x)\}},$$

and for a paired Gehan test,

$$\hat{w}(t, x) = J(t, x)Y_1(t, x)Y_2(t, x)/n_1(t)n_2(t).$$

Variability of these PWLR tests is derived in the Appendix and requires notation for joint and conditional cause-specific hazards of the correlated endpoints. To reduce notation, the pair index, i, will be dropped in the following. Define

$$\begin{aligned}
&\lambda_{g_1 g_2}\{(t_1, x_1), (t_2, x_2)\} \\
&\quad = \lim_{\Delta x_1, \Delta x_2 \to 0} \frac{1}{\Delta x_1 \Delta x_2} \\
&\qquad \times P\{x_1 \le X_{g_1}(t_1) < x_1 + \Delta x_1, \\
&\qquad\quad x_2 \le X_{g_2}(t_2) < x_2 + \Delta x_2, \\
&\qquad\quad \Delta_{g_1}(t_1) = 1, \\
&\qquad\quad \Delta_{g_2}(t_2) = 1 \mid X_{g_1}(t_1) \ge x_1, \\
&\qquad\quad X_{g_2}(t_2) \ge x_2\}
\end{aligned}$$

as the joint cause-specific hazard for the correlated endpoints in groups $g_1 \ne g_2$ at internal times x_1 and x_2, where outcomes related to g_1 and g_2 use data available at calendar times t_1 and t_2, respectively, with $(0 \le x_1 \le t_1, 0 \le x_2 \le t_2)$. Also define the cause specific conditional hazard of failure for treatment group g_1 at study time x_1 as

$$\begin{aligned}
&\lambda_{g_1|g_2}\{(t_1, x_1) \mid (t_2, x_2)\} \\
&\quad = \lim_{\Delta x_1 \to 0} \frac{1}{\Delta x_1} P\{x_1 \le X_{g_1}(t_1) < x_1 + \Delta x_1, \\
&\qquad\quad \Delta_{g_1}(t_1) = 1 \mid X_{g_1}(t_1) \ge x_1, \\
&\qquad\quad X_{g_2}(t_2) \ge x_2\},
\end{aligned}$$

where outcomes related to g_1 and g_2 use data available at calendar times t_1 and t_2, respectively, and where the risk set is restricted to patients with $X_{g_1}(t_1) \ge x_1$ and $X_{g_2}(t_2) \ge x_2$, $(0 \le x_1 \le t_1, 0 \le x_2 \le t_2)$. Let the marginal cause-specific hazard for group g at calendar time t and internal time x, $0 \le x \le t$, be

$$\begin{aligned}
\lambda_g(t, x) = \lim_{\Delta x \to 0} \frac{1}{\Delta x} P\{&x \le X_g(t) < x + \Delta x, \\
&\Delta_g(t) = 1 \mid X_g(t) \ge x\},
\end{aligned}$$

which, under the random censorship assumptions previously described, reduces to the true hazard of T_g, $\lambda_g(x)$, and is not dependent on analysis time t. Let

$$\begin{aligned}
&B_{g_1 g_2}\{(t_1, x_1), (t_2, x_2)\} \\
&\quad = P\{X_{g_1}(t_1) \ge x_1, X_{g_2}(t_2) \ge x_2 \mid E_{g_1} \le t_1, E_{g_2} \le t_2\} \\
&\qquad \times [P\{X_{g_1}(t_1) \ge x_1 \mid E_{g_1} \le t_1\} \\
&\qquad\quad \times P\{X_{g_2}(t_2) \ge x_2 \mid E_{g_2} \le t_2\}]^{-1}.
\end{aligned}$$

Finally, define

$$\begin{aligned}
&G_{g_1 g_2}\{(t_1, x_1), (t_2, x_2)\} \\
&\quad = B_{g_1 g_2}\{(t_1, x_1), (t_2, x_2)\} \\
&\qquad \times [\lambda_{g_1 g_2}\{(t_1, x_1), (t_2, x_2)\} \\
&\qquad\quad - \lambda_{g_1|g_2}\{(t_1, x_1) \mid (t_2, x_2)\}\lambda_{g_2}(x_2) \\
&\qquad\quad - \lambda_{g_2|g_1}\{(t_2, x_2) \mid (t_1, x_1)\}\lambda_{g_1}(x_1) \\
&\qquad\quad + \lambda_{g_1}(x_1)\lambda_{g_2}(x_2)].
\end{aligned}$$

Definitions appropriate for a single analysis use $t_1 = t_2 = t$. Finally, the variance of $\mathcal{T}(t)$, $\sigma^2(t)$, is

$$\sum_{g=1}^{2}\left(\pi_{3-g}(t)\right.$$
$$\left.\times\int_0^\infty [P\{X_g(t)\geq u \mid E_g\leq t\}]^{-1}\{w(t,u)\}^2\lambda_g(u)du\right)$$
$$-\theta(t)\int_0^\infty\int_0^\infty w(t,u)w(t,v)G_{12}\{(t,u),(t,v)\}dvdu,$$

where $\pi_g(t)$ is the probability at calendar time t of being entered into group g with estimate $\hat{\pi}_g(t) = n_g(t)/\{n_1(t) + n_2(t)\}$ and $\theta(t)$ is the proportion of dependent observations in the two treatments at calendar time t with estimate $\hat{\theta}(t) = 2n_{12}(t,t)/\{n_1(t)+n_2(t)\}$. Pooled and unpooled estimates for $\sigma^2(t)$ are given in the Appendix. In cases where individuals enter a study and immediately receive two competing treatments, as in the ETDRS, $\pi_g(t) = 0.5$ and $\theta(t) = 1$. When matched pairs have individual pair members with unentered counterparts at analysis time t, $\theta(t)$ affects the degree to which $\sigma^2(t)$ deviates from the usual variance under independence. When censored time-to-event pairs are independent, $\sigma^2(t)$ corresponds to the variance described by Gill.

Further notation is required to describe the covariance of $\mathcal{T}(t_1)$ and $\mathcal{T}(t_2)$, where $t_1 \leq t_2$. Let

$$H_g(t,x) = P(E_g \leq t - x, V_g \geq x \mid E_g \leq t)$$

be the censoring survival function among those in group g entered by t. Define $\pi_g(t_1 \mid t_2)$ as the probability of entry in group g by t_1 given entry in group g by t_2 with estimate $\hat{\pi}_g(t_1 \mid t_2) = n_g(t_1)/n_g(t_2)$. Let $\theta_{g_1g_2}(t_1,t_2)$ be the proportion of dependent observations in groups g_1 and g_2 at analysis times t_1 and t_2, respectively, with estimate

$$\hat{\theta}_{g_1g_2}(t_1,t_2) = 2n_{g_1g_2}(t_1,t_2)/\{n_{g_1}(t_1) + n_{g_2}(t_2)\}.$$

Let $\gamma_{g_1g_2}(t_1,t_2)$ be the proportion of observations at analysis time t_1 from group g_1 among the total number of observations for group g_1 at time t_1 and for group g_2 at time t_2 with estimate $\hat{\gamma}_{g_1g_2}(t_1,t_2) = n_{g_1}(t_1)/\{n_{g_1}(t_1) + n_{g_2}(t_2)\}$. Define

$$\psi_{g_1g_2}(t_1,t_2)$$
$$= \frac{1}{2}\{\pi_{3-g_1}(t_1)\pi_{3-g_2}(t_2)\}^{\frac{1}{2}}\theta_{g_1g_2}(t_1,t_2)$$
$$\times\left(\left[\frac{\gamma_{g_1g_2}(t_1,t_2)}{1-\gamma_{g_1g_2}(t_1,t_2)}\right]^{\frac{1}{2}} + \left[\frac{\gamma_{g_2g_1}(t_2,t_1)}{1-\gamma_{g_2g_1}(t_2,t_1)}\right]^{\frac{1}{2}}\right).$$

An estimator, $\hat{\psi}_{g_1g_2}(t_1,t_2)$, for $\psi_{g_1g_2}(t_1,\ t_2)$ is constructed from the estimates of its components. Finally, the covariance of $\mathcal{T}(t_1)$ and $\mathcal{T}(t_2)$ is

$$\sigma(t_1,t_2) = \sum_{g=1}^{2}\{\pi_{3-g}(t_1)\pi_{3-g}(t_2)\pi_g(t_1 \mid t_2)\}^{\frac{1}{2}}$$
$$\times\int_0^\infty w(t_1,u)w(t_2,u)$$
$$\times\{S_g(u)H_g(t_2,u)\}^{-1}\lambda_g(u)du$$
$$-\sum_{g=1}^{2}\psi_{g(3-g)}(t_1,t_2)$$
$$\times\int_0^\infty\int_0^\infty w(t_1,u)w(t_2,v)$$
$$\times G_{g(3-g)}\{(t_1,u),(t_2,v)\}dvdu,$$

as shown in the Appendix with pooled and unpooled estimates. If $t_1 = t_2 = t$, $\sigma(t_1,t_2)$ reduces to $\sigma^2(t)$. Otherwise, with dependent endpoints, $\sigma(t_1,t_2)$ does not directly relate to the variance of a single analysis. Dependence in paired censored survival times belies any possibility of an independent increments covariance structure of the repeated tests. This differs from the unpaired log-rank test, which has independent increments. To calculate sequential boundaries, simulation techniques are used. A suitable spending function is selected such as the O'Brien–Fleming (OF) style function, $\alpha_{of} = 2 - 2\Phi(z_{\alpha/2}/v^{1/2})$, where v corresponds to the proportion of information collected at an analysis. Multivariate mean zero normal random variables with the observed covariance of current and previously calculated $\mathcal{T}(t)$ statistics are simulated to estimate critical values giving spending function allocated type I errors. The ETDRS example in Section 3 provides additional instruction on how to construct these boundaries.

3. Simulation Results and ETDRS Example

To verify size of the proposed sequential monitoring strategy, 1000 Monte Carlo simulations with no treatment difference using 150 failure time pairs were generated from the bivariate log-normal distribution for increasing values of correlation. Log scale means and variances were 0.3 and 1, respectively, for each treatment failure time. For each correlated pair, a common uniform$(0,1)$ study entry time was simulated. Paired and unpaired analyses with log-rank and Gehan weights and pooled estimates for variances and covariances were conducted at years 3, 4, and 5 using calendar time as a surrogate for statistical information in the OF spending function with overall type I error of 0.05. Observed sizes located in Table 1 for the unpaired sequentially monitored log-rank test verify the overly conservative nature of analyses that do not take advantage of the correlated failure time structure while PWLR tests give appropriate type I error rates.

Simulations in Table 1 mirroring the above but under an alternative hypothesis with log-scale means of (0.5, 0.3) in the 150 failure time pairs indicate power gains with increasing positive correlation across treatment groups using PWLR tests. In all simulations conducted under the alternative hypothesis, the marginal distributions of the two groups under comparison remain unchanged. Not only do sequential monitoring strategies unadjusted for dependence fail to take advantage of extra precision afforded by the data structure, but power seems to diminish with rising correlation, an effect that can only partially be explained by the observed conservative test sizes. This is likely an artifact of the two estimated weighted cumulative hazards tending to vary in tandem in the presence of positive correlation. The loss of power under comparable marginal distributions using unpaired tests provides further evidence that accounting for the dependent structure of the data is crucial. Table 1 results were essentially unchanged when unpooled variance and covariance estimates were used.

Table 1
Size and power results for paired and unpaired tests[a,b]

		Pair-induced correlation (log scale)			
		0%	30%	60%	90%
Size results	Paired log rank	0.053	0.051	0.055	0.055
	Paired Gehan	0.045	0.047	0.051	0.052
	Log rank	0.052	0.023	0.006	0.000
	Gehan	0.047	0.025	0.002	0.000
Power results	Paired log rank	0.333	0.433	0.643	0.979
	Paired Gehan	0.370	0.483	0.730	0.995
	Log rank	0.344	0.337	0.294	0.176
	Gehan	0.369	0.355	0.329	0.203

[a] 1000 Monte Carlo simulations with 150 censored failure time pairs were generated.
[b] Empirical variance and covariance estimates for the test statistics over 1000 simulations corresponded closely with the average closed-form variance and covariance estimates.

Table 2
ETDRS observed integrated hazard differences and critical values for paired and unpaired analyses

Analysis	Spent error	Observed integrated hazard difference	Paired LR boundary	LR boundary
1	2.85×10^{-5}	0.010	0.024	0.028
2	1.42×10^{-4}	0.014	0.023	0.028
3	5.74×10^{-4}	0.021	0.023	0.029
4	1.18×10^{-3}	0.024	0.025	0.030
5	1.31×10^{-3}	0.022	0.025	0.030
6	2.34×10^{-3}	0.023	0.025	0.031
7	1.33×10^{-3}	0.021	0.025	0.032
⟶ 8	2.27×10^{-3}	0.026	0.025	0.031
9	8.29×10^{-4}	0.027	0.026	0.032

Similar messages appeared in the ETDRS introduced earlier. The DMC, which did not have access to this research, nevertheless recognized statistical issues relating to correlated data. Their exploratory analysis suggested "that not taking pairing into account led to conservative tests" (ETDRS Research Group, 1991a, p. 749). However, their trial was still able to detect a longer time to sight deterioration for the photocoagulation group. To make this example more interesting, it is restricted to 999 patients (1998 eyes) entered prior to February 15, 1983, and taking placebo as part of a separate randomization, reducing the original study size by nearly 75%. The first analysis uses data available on April 8, 1985, when 50 events had been observed, with analyses continued biannually until April 8, 1989. An OF function is used to spend 1% type I error, where the ratio of deaths observed by the interim analysis compared to the total deaths on April 8, 1989, is used as a surrogate for the proportion of information collected.

Table 2 displays the resulting type I errors, observed integrated hazard differences $\{n^*(t)\}^{-1/2}\mathcal{T}$, and estimated critical values corresponding to paired and unpaired log-rank (LR) analyses. In obtaining boundaries, 10,000 multivariate mean zero normal random variables with the observed pooled covariance corresponding to the observed integrated weighted hazard differences were simulated. Specifically, the first cutpoint identifies the value that gives 2.85×10^{-5} type I error in the tails of the first marginal normal distribution. The second cutpoint, which identifies the value giving 1.42×10^{-4} type I error in the tails of the second marginal normal distribution, is estimated among the multivariate normal variates that did not surpass the cutpoint at the first analysis. Using the PWLR tests, a treatment benefit for early photocoagulation is detected at the eighth analysis. Significance would not be achieved without accounting for the correlation in this smaller dataset.

4. Discussion

This research presents closed-form asymptotic distributions of PWLR tests along with nonparametric maximum likelihood-based estimates of relevant variances and covariances and group sequential monitoring procedures related to these statistics. Currently, many trials monitor paired survival endpoints with study designs based on independent samples and accompanying software while acknowledging conservativeness. However, taking advantage of positive correlation in paired outcomes gives large benefits in terms of both type I error and power. Simulations in Section 3 also indicate that, for paired censored survival data alternatives, power using independent

group design and analysis methods might not meet expectations. This is a cause for concern in current practice that the proposed methods eliminate very nicely. Because this work extends a well-understood family of hypothesis tests used in sequential monitoring, the adjusted testing procedures should appeal to the average practitioner since the process of transition to these more efficient tests would be essentially invisible to nonstatistically minded collaborators.

Acknowledgements

The author would like to thank the Early Treatment Diabetic Retinopathy Study Research Group and particularly Marian R. Fisher for the data used in writing this manuscript.

Résumé

Ce travail propose différentes méthodes non paramétriques pour le suivi séquentiel de données de survie appariées, dans le contexte de deux échantillons. Ces méthodes sont basées sur l'utilisation de statistiques du Logrank pondérées adaptées aux données appariées, prenant en compte la dépendance de la durée de survie et de la censure. La forme de la distribution asymptotique jointe de ces statistiques d'analyses séquentielles présente une structure dont les incréments sont dépendants. Des simulations valident les caractéristiques opérationnelles des méthodes proposées, et soulignent les conséquences en terme de puissance et de taille d'échantillon qui résultent de la non prise en compte de données corrélées, même lorsque la dépendance est faible. Une étude sur le traitement précoce des rétinopathies diabétiques permet d'illustrer clairement notre propos.

References

Chang, I., Hsiung, C., and Chuang, Y. (1997). Applications of a frailty model to sequential survival analysis. *Statistica Sinica* **7,** 127–138.

Dabrowska, D. (1989). Rank tests for matched pair experiments with censored data. *Journal of Multivariate Analysis* **28,** 88–114.

Early Treatment Diabetic Retinopathy Study (ETDRS) Research Group. (1991a). Early Treatment Diabetic Retinopathy Study design and baseline patient characteristics: ETDRS report 7. *Ophthalmology* **98,** 741–756.

Early Treatment Diabetic Retinopathy Study (ETDRS) Research Group. (1991b). Early Photocoagulation for Diabetic Retinopathy: ETDRS report 9. *Ophthalmology* **98,** 766–785.

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

Gu, M. G. and Lai, T. L. (1991). Weak convergence of time-sequential censored rank statistics with applications to sequential testing in clinical trials. *Annals of Statistics* **19,** 1403–1433.

Huang, Y. (1999). The two-sample problem with induced dependent censorship. *Biometrics* **55,** 1108–1113.

Lan, K. K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70,** 659–663.

Lin, D. Y. (1991). Nonparametric sequential testing in clinical trials with incomplete multivariate observations. *Biometrika* **78,** 123–131.

Muñoz, S. R., Bangdiwala, S. I., and Sen, P. K. (1997). Group sequential methods for censored bivariate survival data. *Brazilian Journal of Probability and Statistics* **11,** 11–25.

Murray, S. and Cole, B. (2000). Variance and sample size calculations in quality of life adjusted survival analysis (Q-TWiST). *Biometrics* **56,** 266–275.

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics* **35,** 549–556.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64,** 191–199.

Sellke, T. and Siegmund, D. (1983). Sequential analysis of the proportional hazards model. *Biometrika* **70,** 315–326.

Slud, Eric V. (1984). Sequential linear rank tests for two-sample censored survival data. *Annals of Statistics* **12,** 551–571.

Tsiatis, A. A. (1981). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika* **68,** 311–315.

Tsiatis, A. A. (1982). Repeated significance testing for a general class of statistics used in censored survival analysis. *Journal of the American Statistical Association* **77,** 855–861.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84,** 1065–1073.

Received September 1999. *Revised April* 2000. *Accepted May* 2000.

Appendix

Under the null hypothesis, $\mathcal{T}(t)$ is asymptotically equivalent in distribution to

$$\{n^*(t)\}^{1/2} \times \int_0^\infty \hat{w}(t,u) \times [\{Y_1(t,u)\}^{-1} dM_1(t,u) - \{Y_2(t,u)\}^{-1} dM_2(t,u)],$$

where

$$M_g(t,x) = N_g(t,x) - \int_0^x \lambda_g(u) Y_g(t,u) du.$$

For the moment, focus on the term

$$n^{\frac{1}{2}} \int_0^\infty \hat{w}(t,u) \{Y_g(t,u)\}^{-1} dM_g(t,u).$$

After an application of the martingale central limit theorem similar to that used in the Appendix of Murray and Cole (2000), this term is asymptotically equivalent in distribution to

$$Z_g(t) = n^{-\frac{1}{2}} \int_0^\infty w(t,u) [P\{X_g(t) \geq u\}]^{-1} dM_g(t,u),$$

where $w(t,u) = 0$ for $u > t$. Relevant empirical variance and covariance estimators based on $Z_g(t)$ may be constructed using arguments similar to Wei, Lin, and Weissfeld (1989). However, additional work leads to recommended asymptotic

closed forms that clearly demonstrate a dependent increments structure and are estimable using efficient nonparametric maximum likelihood estimates. A result from Gu and Lai (1991) gives that $Z_g(t_1)$ and $Z_g(t_2)$ are asymptotically jointly normal mean zero processes with

$$\begin{aligned}&\operatorname{cov}\{Z_g(t_1), Z_g(t_2)\}\\&\quad = \int_0^\infty w(t_1,u)w(t_2,u)[P\{X_g(t_2) > u\}]^{-1}\lambda_g(u)du.\end{aligned}$$

Since $X_g(t_2) = \min(T_g, V_g, t_2 - E_g)$,

$$\begin{aligned}P\{X_g(t_2) \geq x\} &= P(T_g \geq x, V_g \geq x, t_2 - E_g \geq x)\\&= S_g(x)C_g(t_2,x),\end{aligned}$$

where $S_g(x) = P(T_g \geq x)$ and $C_g(t_2,x) = P(E_g \leq t_2 - x, V_g \geq x)$. Hence,

$$\begin{aligned}&\operatorname{cov}\{Z_g(t_1), Z_g(t_2)\}\\&\quad = \int_0^\infty w(t_1,u)w(t_2,u)\{S_g(u)C_g(t_2,u)\}^{-1}\lambda_g(u)du\\&\quad = \{P(E_g \leq t_2)\}^{-1}\\&\qquad \times \int_0^\infty w(t_1,u)w(t_2,u)\{S_g(u)H_g(t_2,u)\}^{-1}\lambda_g(u)du\end{aligned}$$

asymptotically. For independent treatment groups with $w(t,u) = w(u)$, an independent increments setting would result. However, the joint distribution of $\mathcal{T}(t_1)$ and $\mathcal{T}(t_2)$ also requires $\operatorname{cov}\{Z_{g_1}(t_1), Z_{g_2}(t_2)\}$ for $g_1 \neq g_2$. An application of the multivariate central limit theorem gives

$$\begin{aligned}&\operatorname{cov}\{Z_{g_1}(t_1), Z_{g_2}(t_2)\}\\&\quad = \{P(E_{g_1} \leq t_1)P(E_{g_2} \leq t_2)\}^{-1}P(E_{g_1} \leq t_1, E_{g_2} \leq t_2)\\&\qquad \times \int_0^\infty\int_0^\infty w(t_1,u)w(t_2,v)G_{g_1g_2}\{(t_1,u),(t_2,v)\}dvdu.\end{aligned}$$

Define $\pi_g(t)$ as the probability of entering group g by calendar time t and $\pi_g(t_1 \mid t_2)$ as the probability of entering group g by t_1 given entry in g by t_2. Note that

$$\begin{aligned}&\{n_g(t_2)\}^{-1}\{n^*(t_1)n^*(t_2)\}^{\frac{1}{2}}\\&\qquad \xrightarrow{P} \{\pi_{3-g}(t_1)\pi_{3-g}(t_2)\pi_g(t_1 \mid t_2)\}^{\frac{1}{2}},\end{aligned}$$

$g = 1, 2$. So

$$\begin{aligned}\tau(t_1,t_2) &= \{n^*(t_1)\}^{1/2}\{n^*(t_2)\}^{1/2}n^{-1}\\&\quad \times [\operatorname{cov}\{Z_1(t_1), Z_1(t_2)\} + \operatorname{cov}\{Z_2(t_1), Z_2(t_2)\}\\&\qquad - \operatorname{cov}\{Z_1(t_1), Z_2(t_2)\}\\&\qquad - \operatorname{cov}\{Z_2(t_1), Z_1(t_2)\}]\\&\approx \sum_{g=1}^{2}\{\pi_{3-g}(t_1)\pi_{3-g}(t_2)\pi_g(t_1 \mid t_2)\}^{\frac{1}{2}}\\&\qquad \times \int_0^\infty w(t_1,u)w(t_2,u)\{S_g(u)H_g(t_2,u)\}^{-1}\lambda_g(u)du\\&\quad - \sum_{g=1}^{2}\psi_{g(3-g)}(t_1,t_2)\\&\qquad \times \int_0^\infty\int_0^\infty w(t_1,u)w(t_2,v)G_{g(3-g)}\\&\qquad \times \{(t_1,u),(t_2,v)\}dvdu\end{aligned}$$

becomes the asymptotic covariance for $\mathcal{T}(t_1)$ and $\mathcal{T}(t_2)$. Taking $t_1 = t_2 = t$ provides $\sigma^2(t)$, the variance of $\mathcal{T}(t)$ for a single analysis.

In estimating joint and conditional terms relating to group g_1 at time t_1 and group g_2 at time t_2, attention is restricted to the $n_{g_1g_2}(t_1,t_2)$ correlated pairs where both members entered prior to their respective analysis times. In estimating marginal terms relating to group g at time t, all individual pair members entered in group g by time t will be used regardless of entry by their correlated counterpart. Let

$$\begin{aligned}&Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}\\&\quad = \sum_{k=1}^{n_{g_1g_2}(t_1,t_2)} I\{X_{g_1k}(t_1) \geq x_1, X_{g_2k}(t_2) \geq x_2\}\end{aligned}$$

count correlated pairs where, at analysis time t_1, the group g_1 pair member is still at risk at study time x_1 and, at analysis time t_2, the g_2 pair member is still at risk at study time x_2. Also, let

$$\begin{aligned}&dN_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}\\&\quad = \sum_{k=1}^{n_{g_1g_2}(t_1,t_2)} \lim_{\Delta x_1,\Delta x_2} I\{x_1 \leq X_{g_1k}(t_1) < x_1 + \Delta x_1,\\&\qquad\qquad x_2 \leq X_{g_2k}(t_2) < x_2 + \Delta x_2,\\&\qquad\qquad \Delta_{g_1k}(t_1) = 1, \Delta_{g_2k}(t_2) = 1\}\end{aligned}$$

count correlated pairs where, at analysis time t_1, the g_1 pair member fails at study time x_1 and, at analysis time t_2, the g_2 pair member fails at study time x_2. Let

$$\begin{aligned}&dN_{g_1|g_2}\{(t_1,x_1) \mid (t_2,x_2)\}\\&\quad = \sum_{k=1}^{n_{g_1g_2}(t_1,t_2)} \lim_{\Delta x_1 \to 0} I\{x_1 \leq X_{g_1k}(t_1) < x_1 + \Delta x_1,\\&\qquad\qquad X_{g_2k}(t_2) \geq x_2, \Delta_{g_1k}(t_1) = 1\}\end{aligned}$$

count correlated pairs where, at analysis time t_1, the g_1 pair member had been at risk until failing at study time x_1 and the g_2 pair member at analysis time t_2 remains at risk at study time x_2. An unpooled estimate for

$$G_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}dx_1dx_2$$

becomes

$$\begin{aligned}&\hat{G}_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}\\&\quad = n_{g_1}(t_1)n_{g_2}(t_2)\{n_{g_1g_2}(t_1,t_2)\}^{-1}\\&\qquad \times Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}\{Y_{g_1}(t_1,x_1)Y_{g_2}(t_2,x_2)\}^{-1}\\&\qquad \times ([Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}]^{-1}dN_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}\\&\qquad\quad - [Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}Y_{g_2}(t_2,x_2)]^{-1}\\&\qquad\quad \times dN_{g_1|g_2}\{(t_1,x_1) \mid (t_2,x_2)\}dN_{g_2}(t_2,x_2)\\&\qquad\quad - [Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}Y_{g_1}(t_2,x_1)]^{-1}\\&\qquad\quad \times dN_{g_2|g_1}\{(t_2,x_2) \mid (t_1,x_1)\}dN_{g_1}(t_2,x_1)\\&\qquad\quad + \{Y_{g_1}(t_2,x_1)Y_{g_2}(t_2,x_2)\}^{-1}dN_{g_1}(t_2,x_1)\\&\qquad\quad \times dN_{g_2}(t_2,x_2)).\end{aligned}$$

Hence, an unpooled estimate for $\sigma^2(t)$ is

$$\hat{\sigma}^2(t) = \sum_{g=1}^{2} \hat{\pi}_{3-g}(t) \times \left[\int_0^\infty n_g(t)\{Y_g(t,u)\}^{-2}\{\hat{w}(t,u)\}^2 dN_g(t,u) \right] - \hat{\theta}(t) \int_0^\infty \int_0^\infty \hat{w}(t,u)\hat{w}(t,v)\hat{G}_{12}\{(t,u),(t,v)\}.$$

Under the null hypothesis, some elements of $\sigma^2(t)$ can be estimated by pooling. For times-to-event in groups g_1 and g_2 with entry prior to t, let $\tilde{KM}(t,x)$ be the pooled Kaplan–Meier (KM) estimator for the left-continuous survivor function at study time x. Let $\hat{H}_g(t,x)$ be the KM estimate of the left-continuous censoring survival function for group g at time t. Let $\bar{Y}(t,x) = Y_{g_1}(t,x) + Y_{g_2}(t,x)$ and $\bar{N}(t,x) = N_{g_1}(t,x) + N_{g_2}(t,x)$. A pooled estimate for

$$G_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}dx_1dx_2$$

is

$$\begin{aligned}
&\tilde{G}_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\} \\
&\quad = Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\} \\
&\qquad \times \{n_{g_1g_2}(t_1,t_2)\tilde{KM}(t_1,x_1)\tilde{KM}(t_2,x_2) \\
&\qquad\quad \times \hat{H}_{g_1}(t_1,x_1)\hat{H}_{g_2}(t_2,x_2)\}^{-1} \\
&\qquad \times ([Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}]^{-1} dN_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\} \\
&\qquad\quad - [Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}\bar{Y}(t_2,x_2)]^{-1} \\
&\qquad\quad \times dN_{g_1|g_2}\{(t_1,x_1) \mid (t_2,x_2)\}d\bar{N}(t_2,x_2) \\
&\qquad\quad - [Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}\bar{Y}(t_2,x_1)]^{-1} \\
&\qquad\quad \times dN_{g_2|g_1}\{(t_2,x_2) \mid (t_1,x_1)\}d\bar{N}(t_2,x_1) \\
&\qquad\quad + \{\bar{Y}(t_2,x_1)\bar{Y}(t_2,x_2)\}^{-1} d\bar{N}(t_2,x_1)d\bar{N}(t_2,x_2)).
\end{aligned}$$

So a pooled estimate for $\sigma^2(t)$ is

$$\begin{aligned}
\tilde{\sigma}^2(t) = \sum_{g=1}^{2} \hat{\pi}_{3-g}(t) &\times \left[\int_0^\infty \{\hat{w}(t,u)\}^2 \times \{\hat{H}_g(t,u)\tilde{KM}(t,u)\bar{Y}(t,u)\}^{-1} \times d\bar{N}(t,u) \right] \\
&- \hat{\theta}(t) \int_0^\infty \int_0^\infty \hat{w}(t,u)\hat{w}(t,v)\tilde{G}_{12}\{(t,u),(t,v)\}.
\end{aligned}$$

An unpooled estimate for $\sigma(t_1,t_2)$ is

$$\begin{aligned}
\hat{\sigma}(t_1,t_2) = &\sum_{g=1}^{2} \{\hat{\pi}_{3-g}(t_1)\hat{\pi}_{3-g}(t_2)\hat{\pi}_g(t_1 \mid t_2)\}^{\frac{1}{2}} \\
&\times \int_0^\infty n_g(t_2)\hat{w}(t_1,u)\hat{w}(t_2,u) \\
&\quad \times \{Y_g(t_2,u)\}^{-2} dN_g(t_2,u) \\
&- \sum_{g=1}^{2} \hat{\psi}_{g(3-g)}(t_1,t_2) \\
&\times \int_0^\infty \int_0^\infty \hat{w}(t_1,u)\hat{w}(t_2,v) \\
&\quad \times \hat{G}_{g(3-g)}\{(t_1,u),(t_2,v)\}.
\end{aligned}$$

Pooling under the null hypothesis gives

$$\begin{aligned}
\tilde{\sigma}(t_1,t_2) = &\sum_{g=1}^{2} \{\hat{\pi}_{3-g}(t_1)\hat{\pi}_{3-g}(t_2)\hat{\pi}_g(t_1 \mid t_2)\}^{\frac{1}{2}} \\
&\times \int_0^\infty \hat{w}(t_1,u)\hat{w}(t_2,u) \\
&\quad \times \{\hat{H}_g(t_2,u)\tilde{KM}(t_2,u)\bar{Y}(t_2,u)\}^{-1} \\
&\quad \times d\bar{N}(t_2,u) \\
&- \sum_{g=1}^{2} \hat{\psi}_{g(3-g)}(t_1,t_2) \\
&\times \int_0^\infty \int_0^\infty \hat{w}(t_1,u)\hat{w}(t_2,v) \\
&\quad \times \tilde{G}_{g(3-g)}\{(t_1,u),(t_2,v)\}.
\end{aligned}$$