STATISTICS IN MEDICINE
Statist. Med. 2002; **21**:177–189 (DOI: 10.1002/sim.990)

Group sequential monitoring of years of life saved with paired censored survival data

Susan Murray*,†

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.

SUMMARY

This research develops non-parametric methodology for sequential monitoring of paired time-to-event data when comparing years of life saved, or years where any unpleasant outcome is delayed, is of interest. The recommended family of test statistics uses integrated differences in survival estimates that are available during the study period, where adjustments are made for dependence in the survival and censoring outcomes under comparison. In the context of paired censored survival data, the joint asymptotic closed form distribution of these sequentially monitored test statistics is developed and shown to have a dependent increments structure. Simulations verifying nice operating characteristics of the proposed monitoring methods also reveal consequences of ignoring an underlying paired data structure in terms of size and power properties. A motivating example is also presented via the Early Treatment Diabetic Retinopathy Study, which did not have methods available for sequentially monitoring paired censored survival data at the time. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: clinical trial; correlated times-to-event; Pepe-Fleming statistic; spending function; two-sample test

1. INTRODUCTION

Paired designs have historically been utilized to minimize extraneous sources of variability in making treatment comparisons. Advantages of this study design in positively correlated pairs include increased power to detect treatment differences as compared to similarly sized studies on independent treatment groups. Conversely, if a particular power is desired, paired designs usually achieve the desired power with a smaller required sample size than designs involving independent treatment groups. When paired designs are based on time-to-event endpoints that are available very quickly, tests such as Wilcoxon's signed-rank test or a standard paired t-test may be employed in an analysis. In cases where the endpoint of interest takes longer to observe, counterparts to the paired t-test that can accommodate right censoring are desirable.

A counterpart to the two-sample t-test for independent groups that has received attention in the censored data setting is the years of life saved (YLS) test, which has been independently developed for the group sequential setting by Murray and Tsiatis [1] and Li [2], after its initial discussion by Pepe and Fleming [3]. These tests compare integrated Kaplan–Meier survival

*Correspondence to: Susan Murray, Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109, U.S.A.
†E-mail: skmurray@umich.edu

Received June 2000
Accepted March 2001

estimates, sometimes weighted, during the study period. Murray [4] recently extended YLS tests to accommodate paired censored survival data in the case of a single analysis. Since integrated Kaplan–Meier curves correspond to the sample mean in the absence of censoring, YLS tests constructed with correlated Kaplan–Meier curves are the closest relatives to the paired t-test in the censored survival setting. YLS tests have been frequently lauded as an alternative to the logrank (LR) test, particularly attractive when hazards are not proportional in nature, and are straightforward to communicate to audiences with an emphasis in clinical rather than statistical training. For instance, in collaborating with diabetic retinopathy investigators in studying paired censored time to vision loss, YLS methods provide point estimates and confidence intervals on the average extended time of vision for eyes on the superior intervention during the study period.

The Early Treatment Diabetic Retinopathy Study (ETDRS) Research Group collected information on severe vision loss in patients suited to this purpose, where severe visual loss was defined as visual acuity less than 5/200 at two consecutive visits. This group enrolled 3711 patients with mild-to-severe non-proliferative or early proliferative diabetic retinopathy in both eyes from April 1980 to July 1985 [5, 6]. One eye of each patient was randomized to early photocoagulation and the other to deferral of photocoagulation until such time when high-risk proliferative retinopathy was detected. Because patients were recruited and followed in the ETDRS over a period of nine years, ethical considerations required periodic monitoring of accumulating time-to-event data in order to ensure timely detection of treatment benefits or detriments among the study participants. The ETDRS therefore would prepare reports at least twice a year for a Data Monitoring Committee that would subsequently use the information in determining whether the trial should end early or be continued. Hence appropriate YLS analysis methods require correct handling of censoring, pairing and sequential monitoring issues with respect to the data. Principles of group sequential monitoring outlined by Pocock [7], O'Brien and Fleming [8] and Lan and DeMets [9] set the standard for extending YLS methods to paired censored survival data monitored in the ETDRS trial.

So far little group sequential methodology has been developed for use in the paired censored survival data setting as in the ETDRS. Chang *et al.* [10] considered sequential methods for frailty models assuming common pair entry times. Murray [11] developed methods for group sequential monitoring of paired weighted LR tests that allowed entry times to vary within the correlated pairs. Other authors have studied sequential designs for independent treatment groups where multiple correlated censored survival outcomes are monitored. For instance, Lin [12] devised a non-parametric weighted linear rank statistic for monitoring correlated non-identically distributed outcome types subject to censoring across two independent groups, while Muñoz *et al.* [13] proposed parametric models for sequentially monitoring correlated pairs of similar outcome types subject to censoring across independent groups. However, group sequential development of the YLS method for matched pair experiments with censored data is currently unavailable for studies designed similarly to the ETDRS.

This research makes available methods for sequential analysis of paired YLS tests. As in all group sequential methods, the key to the sequential monitoring lies in understanding the joint distribution of the repeated statistical tests. Section 2 describes the paired YLS test in the case where a single analysis takes place and the joint distribution of these tests at different analysis times. Results from this section reveal that the dependent nature of the paired outcomes belies any possibility of an independent increments covariance structure of the repeated tests. Indications for how to use the joint distribution to produce stopping boundaries are also

given in this section. Simulations verifying the operating characteristics of the recommended sequential monitoring procedure are given in Section 3. Comparisons are made with sequential monitoring procedures that do not account for the dependent structure of the paired censored time-to-event outcomes. In Section 4, an example relating to the ETDRS study is given. A discussion follows in Section 5.

2. JOINT DISTRIBUTION OF PAIRED YLS STATISTICS

To understand concepts relating to sequential theory in the setting of dependent time-to-event outcomes being compared across time, 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 previously mentioned ETDRS example, or potentially a matched pair whose members are randomized to receive opposing treatment regimens, as in a study on siblings. These n individuals or n matched pairs enter the trial at times E_{gi}, for $i=1,\ldots,n$ and $g=1,2$ during the accrual period. In many cases $E_{1i}=E_{2i}$, referring to the single entry time of an individual denoted by i, otherwise E_{1i} and E_{2i} denote potentially different entry times. Entry times are assumed to be identically distributed within treatment group g with $E_{g_1 i_1}$ independent of $E_{g_2 i_2}$ for $i_1 \neq i_2$. Each individual or matched pair denoted by i has two correlated survival times T_{gi}, $g=1,2$ measured from the time of entry. For instance in the ETDRS, T_{1i} and T_{2i} measure time from randomization to an objective measure of severe vision loss for eyes randomized to deferred and early photocoagulation, respectively, within an individual. Time-to-event outcomes for each individual or matched pair that have not occurred prior to the time of analysis are censored. For instance, if the data were analysed at calendar time t, one would censor outcomes where $T_{gi} > t - E_{gi}$. The notation V_{gi}, $g=1,2$, $i=1,\ldots,n$, will be used to refer to the potential censoring times due to random loss to follow-up. Aside from potential dependence allowed 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 $i=1,\ldots,n$. If the data were analysed at calendar time t, then the observable random variables for treatment group g would be $\{X_{gi}(t), \Delta_{gi}(t)\}$, for all $i=1,\ldots,n$ such that $E_{gi} \leqslant t$, where $X_{gi}(t)=\min(T_{gi}, V_{gi}, t-E_{gi})$ is the observed time on study at analysis time t and $\Delta_{gi}(t)=I\{T_{gi} \leqslant \min(t-E_{gi}, V_{gi})\}$ denotes the failure indicator at calendar time t. Indices referring to calendar time measured from the start of the study, and indices referring to time measured from a patient's entry into the study will frequently be used in combination. Hence note that the index 't' will index calendar time of an analysis, and the index 'x' will index time from entry into the study, commonly referred to as study time.

Define the total sample size enrolled at calendar time t in group g as $n_g(t)=\sum_{i=1}^{n} I(E_{gi} \leqslant t)$. In order to keep track of the number of correlated entered pairs across treatment groups g_1, g_2 and calendar times t_1, t_2 we define $n_{g_1 g_2}(t_1,t_2)=\sum_{i=1}^{n} I(E_{g_1 i} \leqslant t_1, E_{g_2 i} \leqslant t_2)$. Note that when pairs of dependent outcomes are attributed to an individual, one will often have $n_1(t)=n_2(t)=n_{12}(t,t)$. If at the final analysis time all treatment pairs have been entered, one will have $n_1(t)=n_2(t)=n_{12}(t,t)=n$. However this method also allows for the case where some individual pair members remain unentered at the time of the final analysis, as long as the number of complete pairs entered is approaching infinity.

For each treatment group g and calendar analysis time t, define the number of individuals at risk at study time x as $Y_g(t,x)=\sum_{i=1}^{n} I\{X_{gi}(t) \geqslant x\}$ and let $J(t,x)=1$ if $Y_1(t,x)Y_2(t,x)>0$ and

$J(t,x)=0$ otherwise. Let $n^*(t)=n_1(t)n_2(t)/\{n_1(t)+n_2(t)\}$. At each analysis time t, consider a paired YLS test

$$\mathscr{T}(t)=\{n^*(t)\}^{\frac{1}{2}}\int_0^\infty J(t,u)\{\hat{S}_1(t,u)-\hat{S}_2(t,u)\}\,\mathrm{d}u \tag{1}$$

where $\hat{S}_g(t,x)$ is the Kaplan–Meier survival estimate for the true survival at study time x, $S_g(x)$, using information available for all individuals entered in group g at analysis time t, regardless of whether their correlated counterpart has entered the study or not. The remainder of this section describes results on variability of this statistic when it is sequentially monitored and methods for finding statistical significance cutpoints that will protect the overall type I error rate of the trial. Some readers may choose to avoid notational definitions relating to the covariance structure of the sequentially monitored statistics by skipping ahead to the last paragraph of this section.

The variability of this statistic at a single analysis time requires notation relating to joint and conditional hazards of the correlated endpoints. It will later be convenient to allow for the correlated random variables of interest in the following definitions to have different amounts of calendar time follow-up according to their use at analysis times t_1 and t_2. Definitions appropriate for a single analysis time would use $t_1=t_2=t$. Also, to reduce notation, the index i referring to the individual or matched pair will be suppressed in the following definitions of joint and conditional terms. Define $\lambda_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}=\lim_{\Delta x_1,\Delta x_2\to 0}P(x_1\leqslant X_{g_1}(t_1)<x_1+\Delta x_1,\ x_2\leqslant X_{g_2}(t_2)<x_2+\Delta x_2,\ \Delta_{g_1}(t_1)=1,\ \Delta_{g_2}(t_2)=1|X_{g_1}(t_1)\geqslant x_1,\ X_{g_2}(t_2)\geqslant x_2)/(\Delta x_1\Delta x_2)$ to be a joint hazard function for the correlated endpoints in treatment groups $g_1\neq g_2$ at study times x_1 and x_2 where outcomes related to g_1 are subject to data available at calendar time t_1 and outcomes related to g_2 are subject to data available at calendar time t_2, $(0\leqslant x_1\leqslant t_1, 0\leqslant x_2\leqslant t_2)$. Also define the conditional hazard function $\lambda_{g_1|g_2}\{(t_1,x_1)|(t_2,x_2)\}=\lim_{\Delta x_1\to 0}P(x_1\leqslant X_{g_1}(t_1)<x_1+\Delta x_1,\Delta_{g_1}(t_1)=1|X_{g_1}(t_1)\geqslant x_1,X_{g_2}(t_2)\geqslant x_2)/\Delta x_1$, which may be interpreted as the hazard of failure for treatment group g_1 at study time x_1 where again outcomes related to g_1 are subject to data available at calendar time t_1 and outcomes related to g_2 are subject to data available at calendar time t_2 and where the risk set is restricted to those patients with $X_{g_1}(t_1)\geqslant x_1$ and $X_{g_2}(t_2)\geqslant x_2$, $(0\leqslant x_1\leqslant t_1, 0\leqslant x_2\leqslant t_2)$. Also define the marginal hazard for treatment group g at calendar time t and study time x, $0\leqslant x\leqslant t$, to be $\lambda_g(t,x)=\lim_{\Delta x\to 0}P(x\leqslant X_g(t)<x+\Delta x,\Delta_g(t)=1|X_g(t)\geqslant x)/(\Delta x)$, 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

$$B_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}=\frac{P(X_{g_1}(t_1)\geqslant x_1,X_{g_2}(t_2)\geqslant x_2|E_{g_1}\leqslant t_1,E_{g_2}\leqslant t_2)}{P(X_{g_1}(t_1)\geqslant x_1|E_{g_1}\leqslant t_1)P(X_{g_2}(t_2)\geqslant x_2|E_{g_2}\leqslant t_2)}$$

Define

$$\begin{aligned}G_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}=B_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}[&\lambda_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}\\&-\lambda_{g_1|g_2}\{(t_1,x_1)|(t_2,x_2)\}\lambda_{g_2}(x_2)-\lambda_{g_2|g_1}\{(t_2,x_2)|(t_1,x_1)\}\lambda_{g_1}(x_1)\\&+\lambda_{g_1}(x_1)\lambda_{g_2}(x_2)]\end{aligned}$$

Also, to reduce notation, define $A(t,x)=\int_x^\infty p(t,u)S(u)\,\mathrm{d}u$, where $p(t,x)=I[P\{X_1(t)\geqslant x\}\times P\{X_2(t)\geqslant x\}>0]$. In the case where a single analysis occurs at analysis time t, the

asymptotic variance of $\mathscr{T}(t)$ is

$$\sigma^2(t) = \sum_{g=1}^{2} \frac{\pi_1(t)\pi_2(t)}{\pi_g(t)} \left[\int_0^\infty \frac{\{A_g(t,u)\}^2 \lambda_g(u)}{P(X_g(t) \geqslant u | E_g \leqslant t)} \, \mathrm{d}u \right]$$

$$-\theta(t) \int_0^\infty \int_0^\infty A_1(t,u) A_2(t,v) G_{12}\{(t,u),(t,v)\} \, \mathrm{d}v \, \mathrm{d}u$$

where $\pi_g(t) = \lim_{\{n_1(t), n_2(t) \to \infty\}} n_g(t)/\{n_1(t) + n_2(t)\}$ is the probability at calendar time t of being entered into treatment group g with estimate $\hat{\pi}_g(t) = n_g(t)/\{n_1(t) + n_2(t)\}$ and $\theta(t) = \lim_{\{n_{12}(t,t) \to \infty\}} 2n_{12}(t,t)/\{n_1(t) + n_2(t)\}$ is the sampling proportion of dependent observations in the two treatment groups at calendar time t with estimate $\hat{\theta}(t) = 2n_{12}(t,t)/\{n_1(t) + n_2(t)\}$. Pooled and unpooled estimation procedures for $\sigma^2(t)$ are given in the Appendix. In many cases where individuals are entered into the study and immediately given the two competing treatments as in the ETDRS example, $\pi_g(t) = 0.5$ and $\theta(t) = 1$. In cases where matched pairs have different random entry times into the trial there may be a subset of individual pair members with an unentered counterpart at the time of analysis t and $\theta(t)$ can be interpreted as affecting the degree to which the final term of $\sigma^2(t)$ deviates from the usual variance under independent treatment groups. Also note that when censored time-to-event pairs are truly independent in nature, $\sigma^2(t)$ will correspond to the usual variance described by Pepe and Fleming under independent treatment groups.

Further notation is required to describe the covariance of $\mathscr{T}(t_1)$ and $\mathscr{T}(t_2)$, where without loss of generality this paper will assume $t_1 \leqslant t_2$. Let $H_g(t,x) = P(E_g \leqslant t - x, V_g \geqslant x | E_g \leqslant t)$ be the censoring survival distribution among individuals in treatment group g entered by calendar time t. Define $\pi_g(t_1|t_2) = \lim_{\{n_g(t_1) \to \infty\}} n_g(t_1)/n_g(t_2)$ as the probability of entry in group g by t_1 given entry in group g by t_2 with estimate $\hat{\pi}_g(t_1|t_2) = n_g(t_1)/n_g(t_2)$. Let $\theta_{g_1g_2}(t_1,t_2) = \lim_{\{n_{g_1g_2}(t_1,t_2) \to \infty\}} 2n_{g_1g_2}(t_1,t_2)/\{n_{g_1}(t_1) + n_{g_2}(t_2)\}$ be the sampling proportion of dependent observations in treatment group g_1 at analysis time t_1 and treatment group g_2 at analysis time t_2 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 $\gamma_{g_1g_2}(t_1,t_2) = \lim_{\{n_{g1}(t_1), n_{g2}(t_2) \to \infty\}} n_{g_1}(t_1)/\{n_{g_1}(t_1) + n_{g_2}(t_2)\}$ be the sampling proportion of observations available at analysis time t_1 from treatment group g_1 among the total number of observations available for treatment group g_1 at analysis time t_1 and for treatment group g_2 at analysis 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 $\psi_{g_1g_2}(t_1,t_2) = \frac{1}{2}\{\pi_{3-g_1}(t_1)\pi_{3-g_2}(t_2)\}^{1/2} \theta_{g_1g_2}(t_1,t_2)([\gamma_{g_1g_2}(t_1,t_2)/\{1 - \gamma_{g_1g_2}(t_1,t_2)\}]^{1/2} + [\gamma_{g_2g_1}(t_2,t_1)/\{1 - \gamma_{g_2g_1}(t_2,t_1)\}]^{1/2})$. An estimator, $\hat{\psi}_{g_1g_2}(t_1,t_2)$, for $\psi_{g_1g_2}(t_1,t_2)$ is constructed easily from the estimates of its components.

Finally, the covariance of $\mathscr{T}(t_1)$ and $\mathscr{T}(t_2)$ is

$$\sigma(t_1,t_2) = \sum_{g=1}^{2} \{\pi_{3-g}(t_1)\pi_{3-g}(t_2)\pi_g(t_1|t_2)\}^{1/2}$$

$$\times \int_0^\infty A_g(t_1,u) A_g(t_2,u) \{S_g(u) H_g(t_2,u)\}^{-1} \lambda_g(u) \, \mathrm{d}u$$

$$-\sum_{g=1}^{2} \psi_{g(3-g)}(t_1,t_2) \int_0^\infty \int_0^\infty A_g(t_1,u) A_{3-g}(t_2,v) G_{g(3-g)}\{(t_1,u),(t_2,v)\} \, \mathrm{d}v \, \mathrm{d}u$$

as shown in the Appendix where pooled and unpooled estimates for $\sigma(t_1, t_2)$ are also located. Note that when $n_g(t_1) = n_g(t_2) = n_{gg}(t_1, t_2)$, $\psi_{gg}(t_1, t_2)$ reduces to $\pi_{3-g}(t_1) = \pi_{3-g}(t_2)$ and $\pi_g(t_1|t_2)$ becomes one. If in addition $t_1 = t_2 = t$, $\sigma(t_1, t_2)$ reduces to $\sigma^2(t)$. In general the covariance between $\mathscr{T}(t_1)$ and $\mathscr{T}(t_2)$ does not reduce to the variance at the earlier interim analysis. The joint distribution of $\mathscr{T}(t_1)$ and $\mathscr{T}(t_2)$ does not have an independent increments structure. This differs somewhat from the independent treatment group case, where defining $J(t,x)$ terms to be equivalent at all analysis times t would result in an independent increments structure.

To calculate sequential boundaries in this non-independent increments case, one may use Monte Carlo numerical integration techniques in relation to the joint distribution of the test statistics at the various analysis times. First a suitable spending function is selected such as the O'Brien–Fleming (OF) styled spending function $\alpha_{\mathrm{OF}}(v_j) = 2 - 2\Phi(z_{\alpha/2}/\sqrt{v_j})$, where v_j corresponds to some surrogate for the proportion of information collected at the jth interim analysis time. The surrogate for the proportion of information collected may be chosen to reflect the percentage of calendar time elapsed toward the planned length of the study or may be chosen as the percentage of observed events at the analysis time of those required by the end of the study to achieve the designed power. Both of these choices would be known at the design stage of a clinical trial and either would be appropriate in terms of protecting type I error of the trial. Next the covariance structure between the current and all previous $\mathscr{T}(t)$ test statistics calculated during the course of the trial using the observed data is estimated. At the jth analysis time boundary cutpoints, $c_1, \ldots, c_j$, must be chosen so that

$$\begin{aligned} &P(|\mathscr{T}(t_1)| < c_1, \ldots, |\mathscr{T}(t_{j-1})| < c_{j-1}, |\mathscr{T}(t_j)| > c_j) \\ &\quad = P(|\mathscr{T}(t_j)| > c_j \,\big|\, |\mathscr{T}(t_1)| < c_1, \ldots, |\mathscr{T}(t_{j-1})| < c_{j-1}) P(|\mathscr{T}(t_1)| < c_1, \ldots, |\mathscr{T}(t_{j-1})| < c_{j-1}) \\ &\quad = \alpha_{\mathrm{OF}}(v_j) - \alpha_{\mathrm{OF}}(v_{j-1}) \end{aligned}$$

the type I error to be spent at the jth analysis time. Multivariate mean zero normal random variables with the observed covariance structure are simulated and used to estimate appropriate cut-off points for the statistics at the different analysis times. The process is recursive in nature. For instance after determining $c_1, \ldots, c_{j-1}$, one may easily use the multivariate normal replicates to estimate $P(|\mathscr{T}(t_1)| < c_1, \ldots, |\mathscr{T}(t_{j-1})| < c_{j-1})$ and the boundary c_j may be found by considering the tails of the marginal distribution corresponding to the jth analysis time among multivariate replicates that did not surpass cutpoints at previous analysis times. An example using ETDRS data in Section 4 provides additional instruction on how these simulated boundaries are constructed.

3. SIMULATION RESULTS

In order to verify size of the proposed sequential monitoring strategy, 1000 Monte Carlo simulations under the null hypothesis of no treatment difference were conducted using 150 failure time pairs generated from the bivariate log-normal distribution, where correlations examined between the two failure times on the log scale were {0 per cent, 30 per cent, 60 per cent, 90 per cent}. Log scale means and variances were 0.3 and 1, respectively, for each of the two treatment group failure times. Common and independent entry-time scenarios

Table I. Size and power results for paired and unpaired tests.

		Entry-time correlation	Event-time correlation (log scale) 0%	30%	60%	90%
Size results	Paired YLS	1	0.046	0.045	0.048	0.040
		0	0.055	0.043	0.039	0.046
	YLS	1	0.043	0.026	0.005	0.000
		0	0.057	0.022	0.003	0.000
Power results	Paired YLS	1	0.361	0.464	0.691	0.995
		0	0.373	0.473	0.663	0.997
	YLS	1	0.368	0.329	0.321	0.179
		0	0.375	0.322	0.314	0.172

A total of 1000 Monte Carlo simulations with 150 censored failure time pairs were generated.
Empirical variance and covariance estimates for the test statistics over 1000 simulations corresponded closely with the average closed form variance and covariance estimates.

for pair members were simulated using the Uniform(0, 1) distribution. Interim analyses were conducted at years 3, 4 and 5 using calendar time as a surrogate for statistical information in the O'Brien–Fleming spending function with an overall type I error of 0.05. Observed paired and unpaired YLS test sizes in increasing order of correlation using pooled estimates for variances and covariances are located in the upper panel of Table I. Similar results were observed using unpooled variance and covariance estimates. Sizes corresponding to the paired tests have appropriate type I error levels for all degrees of correlation, while unpaired tests become increasingly conservative as underlying correlation in the survival times grows. Entry-time correlation appears to have little bearing on the performance of the sequentially monitored tests.

To study operating characteristics under an alternative hypothesis, 150 failure time pairs were generated from the bivariate log-normal distribution with log scale means of (0.5, 0.3) with variability parameters unchanged from the simulations described above. Results for paired and unpaired tests are located in the lower panel of Table I using the O'Brien–Fleming spending function and pooled variance and covariance estimates. Note that in all simulations conducted under the alternative hypothesis, the marginal distributions of the two groups under comparison remain unchanged. Monitoring strategies that take into account the paired correlation structure increase in power with growing correlation in survival times for comparable group marginal distributions. In contrast, the usual sequential monitoring strategy that does not take into account dependence between paired survival outcomes loses substantial power as correlation in paired survival times increases. These unattractive power and size results related to unpaired tests are likely an artefact of the two estimated survival curves tending to vary in tandem in the presence of positive correlation and give evidence that accounting for the dependent structure of the data is a crucial step in the group sequential analysis.

4. EXAMPLE

Recall the Early Treatment Diabetic Retinopathy Study (ETDRS) described in the introduction, which enrolled 3711 patients with mild-to-severe non-proliferative or early proliferative

Table II. ETDRS observed YLS and integrated hazard differences with corresponding critical values for paired and unpaired analyses.

Analysis	Spent error	Observed integrated hazard difference	Paired LR boundary	LR boundary	Observed YLS	Paired YLS boundary	YLS boundary
1	2.85×10^{-5}	0.010	0.024	0.028	14.70	31.35	34.01
2	1.42×10^{-4}	0.014	0.023	0.028	21.42	30.04	41.11
3	5.74×10^{-4}	0.021	0.023	0.029	31.46	36.00	44.11
4	1.18×10^{-3}	0.024	0.025	0.030	33.56	38.25	45.39
5	1.31×10^{-3}	0.022	0.025	0.030	31.65	39.16	49.50
6	2.34×10^{-3}	0.023	0.025	0.031	38.14	41.21	53.03
7	1.33×10^{-3}	0.021	0.025	0.032	41.34	42.86	54.79
8*	2.27×10^{-3}	0.026	0.025	0.031	52.95	47.06	57.94
9	8.29×10^{-4}	0.027	0.026	0.032	59.92	51.86	64.56

*Null hypothesis rejected using paired boundaries at this analysis time.

diabetic retinopathy in both eyes and randomized one eye of each patient to early photocoagulation and the other to deferral of photocoagulation until a later time when high-risk proliferative retinopathy was detected. The major endpoint of interest was time to severe visual loss. The Data Monitoring Committee prepared interim reports approximately twice a year using statistical methods of comparison which did not take into account the pairing of eyes on study. The first interim analysis took place when 50 events had occurred across the two comparison groups. A statistically significant result during monitoring was defined as a test statistic with corresponding p-value less than 0.01. The Data Monitoring Committee, which did not have access to methodology for sequentially monitoring paired censored survival data, nevertheless recognized the statistical issues relating to the correlated structure of the data. Some exploratory analysis on their part suggested 'that not taking pairing into account led to conservative tests (reference [5], p. 749)'. However, with their large sample size they were still able to detect a longer time to sight deterioration with early photocoagulation.

As an example of the proposed analysis methods which take into account the natural pairing of the data, the ETDRS study is revisited. To make this example more interesting, the analysis is restricted to those 999 patients (1998 eyes) who entered the study prior to 15 February 1983 and who were simultaneously taking a placebo pill as part of a separate randomization, reducing by nearly 75 per cent the sample size of the original ETDRS study. Following the example of the ETDRS study, the first analysis will take place using data that would have been available on 8 April 1985, when 50 events had been observed, and continue twice a year until 8 April 1989 for a total of nine interim analyses. However, this example will use a more conservative type I error than the original study. After nine analyses, the following strategy will have merely a 1 per cent type I error overall instead of the originally planned 1 per cent error per interim analysis. An O'Brien–Fleming spending function will be employed where the ratio of deaths observed by the interim analysis compared to the total deaths on 8 April 1989 is used as a surrogate for the proportion of information collected. Note that in a prospective clinical trial, one would estimate the total number of deaths required for a well powered design at the last analysis time in this computation.

Results in Table II include interim analysis number, type I error spent at each analysis time, observed integrated hazard differences corresponding to a LR test analysis approach, the

average extended days of sight observed in the early photocoagulation treatment arm during the study period ($\{n^*(t)\}^{-1/2}\mathscr{T}$) and sequential boundaries corresponding to paired and unpaired YLS and LR tests at the nine analysis times. Paired and unpaired LR boundaries were calculated as in Murray [11]. To determine appropriate sequential boundary cutpoints for the YLS analyses, Monte Carlo numerical integration strategies were employed in relation to the joint distribution of the test statistics at the nine interim analyses. First a covariance matrix corresponding to these observed integrated survival differences was constructed using pooled formulae outlined in the Appendix. Multivariate mean zero normal random variables with the observed covariance structure were then simulated. At the jth analysis time, $j=1,\ldots,9$, boundary cutpoints $c_1,\ldots,c_j$ were chosen so that $P(|\mathscr{T}(t_1)|<c_1,\ldots,|\mathscr{T}(t_{j-1})|<c_{j-1},|\mathscr{T}(t_j)|>c_j)$ was equal to the type I error to be spent at the jth analysis time. Specifically, the first cutpoint, c_1, identifies the value which gives 2.85×10^{-5} type I error in the tails of the marginal normal distribution corresponding to the first analysis time, so that $P(|\mathscr{T}(t_1)|>c_1)=2.85\times10^{-5}$. The Monte Carlo results gave $c_1=31.35$ when monitoring with the paired YLS statistic. The second cutpoint, c_2, was chosen so that $P(|\mathscr{T}(t_1)|<31.35,|\mathscr{T}(t_2)|>c_2)=P(|\mathscr{T}(t_2)|>c_2\,\big|\,|\mathscr{T}(t_1)|<31.35)P(|\mathscr{T}(t_1)|<31.35)$ was equal to 1.42×10^{-4}. Hence the value of $c_2=30.04$ was found by considering the tails of the marginal normal distribution corresponding to the second analysis time in multivariate normal replicates that did not surpass the first cutpoint at the first analysis time in combination with the Monte Carlo estimate of $P(|\mathscr{T}(t_1)|<31.35)$. Similarly the value of $c_3=36.00$ was determined by considering the tails of the marginal normal distribution corresponding to the third analysis time among multivariate normal replicates that did not surpass cutpoints at previous analysis times in combination with the Monte Carlo estimate of $P(|\mathscr{T}(t_1)|<31.35,|\mathscr{T}(t_2)|<30.04)$, and so on. In studying the various sequential boundaries in Table II, the null hypothesis was rejected at the 0.01 level at the eighth analysis time using either the paired YLS or the paired LR test. At this analysis 52.95 extra days of sight were observed on average in the first 8.47 years of observation in the early photocoagulation group (95 per cent confidence interval 19.36, 86.54 extra days of sight), although this estimate and corresponding confidence interval are slightly inflated due to the nature of the sequential stopping rule. Neither of the monitoring strategies that ignore the correlated nature of the pairs was able to achieve statistical significance in this smaller data set.

5. DISCUSSION

This research presents closed form asymptotic distributions of years of life saved tests for use with paired censored survival data and makes available new group sequential monitoring procedures related to these statistics. In studying the joint structure of the recommended test statistics computed over time, asymptotic closed form variances and covariances of the test statistics are provided. Based on these closed form quantities, pooled and unpooled variance and covariance estimates are proposed that, in combination with the recommended monitoring procedure, perform very well towards the goal of protecting the overall type I error whether one or multiple analyses are performed.

Methods in this research can also easily accommodate tests based on integrated weighted differences in survival by including these weighting functions within the $J(t,x)$ component of the tests described herein. Theoretical development remains the same as long as the weighting function converges in probability at all study times, x. If weighting is desired to capture early survival differences with higher probability, for instance, a fixed weighting function with higher

weights at the earlier study times could capture this effectively. Users are cautioned about selecting weights that depend on the censoring mechanism, since at each analysis time the degree of censoring will change and affect the interpretation of the test statistic. This would result in the trial design depending on changing alternatives across analysis times, t. The interpretation issue cited here does not play a role in the sequential analyses of weighted LR style tests, since weights in that setting are designed to emphasize areas where proportional hazards are measured more accurately. Hence the interpretation of sequentially computed weighted logrank statistics remains similar in the case of proportional hazards as the weights change from analysis to analysis.

Currently many clinical trials that monitor paired survival endpoints, such as the ETDRS, employ study designs based on independent samples. Also, because methods based on independent samples are readily available for monitoring purposes, the temptation is to use already available methods while acknowledging their conservativeness. This research demonstrates that taking advantage of the positive correlation structure in the 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 structures, power under the alternative hypothesis might not even match the power aimed for in design when methods for independent samples are used both in design and analysis stages of clinical research. This is a cause for concern in current practice with this data structure that the proposed methods eliminate very nicely.

APPENDIX

A1. Covariance of $\mathscr{T}(t_1)$ and $\mathscr{T}(t_2)$

For each treatment group g and calendar analysis time t, define the number of events occurring no later than study time x as $N_g(t,x) = \sum_{i=1}^{n} I\{X_{gi}(t) \leqslant x, \Delta_{gi}(t) = 1\}$ for $0 \leqslant x \leqslant t$ and let $M_g(t,x) = N_g(t,x) - \int_0^x \lambda_g(u) Y_g(t,u)\,\mathrm{d}u$. Consider $\mathscr{T}(t)$, which after an application of the martingale central limit theorem is asymptotically equivalent in distribution to $Z_1(t) - Z_2(t)$ under the null hypothesis of no treatment difference, where

$$Z_g(t) = \{n^*(t)\}^{1/2} \int_0^\infty p(t,u) S_g(u) \int_0^u [nP\{X_g(t) \geqslant x\}]^{-1}\,\mathrm{d}M_g(t,x)\,\mathrm{d}u.$$

The covariance of interest, $\mathrm{cov}\{\mathscr{T}(t_1), \mathscr{T}(t_2)\}$, becomes

$$\sum\nolimits_{g=1}^{2} \mathrm{cov}\{Z_g(t_1), Z_g(t_2)\} - \sum\nolimits_{g=1}^{2} \mathrm{cov}\{Z_g(t_1), Z_{3-g}(t_2)\}.$$

A result from Murray and Tsiatis [1] gives that $Z_g(t_1)$ and $Z_g(t_2)$ are asymptotically jointly normal mean zero random variables with $\mathrm{cov}\{Z_g(t_1), Z_g(t_2)\} = \{\pi_{3-g}(t_1)\pi_{3-g}(t_2)\pi_g(t_1|t_2)\}^{1/2} \times \int_0^\infty A_g(t_1,u) A_g(t_2,u) \{S_g(u) H_g(t_2,u)\}^{-1} \lambda_g(u)\,\mathrm{d}u$. If our treatment groups were independent, then this last result would give us all needed information to identify the joint distribution of $\mathscr{T}(t_1)$ and $\mathscr{T}(t_2)$. In fact, under the assumption of independent treatment groups and $J(t_1,u) = J(t_2,u)$ for all t_1, t_2 one would have an independent increments setting. However, one also needs

to identify $\mathrm{cov}\{Z_{g_1}(t_1), Z_{g_2}(t_2)\}$ for $g_1 \neq g_2$. In this case an application of the multivariate central limit theorem gives the result

$$\mathrm{cov}\{Z_{g_1}(t_1), Z_{g_2}(t_2)\} = \psi_{g_1g_2}(t_1, t_2) \int_0^\infty \int_0^\infty A_{g_1}(t_1, u) A_{g_2}(t_2, v) G_{g_1g_2}\{(t_1, u), (t_2, v)\}\, \mathrm{d}v\, \mathrm{d}u.$$

So

$$\begin{aligned}\sigma(t_1, t_2) \approx & \sum\nolimits_{g=1}^{2} \{\pi_{3-g}(t_1)\pi_{3-g}(t_2)\pi_g(t_1|t_2)\}^{1/2} \\ & \times \int_0^\infty A_g(t_1, u) A_g(t_2, u)\{S_g(u)H_g(t_2, u)\}^{-1}\lambda_g(u)\, \mathrm{d}u \\ & - \sum\nolimits_{g=1}^{2} \psi_{g(3-g)}(t_1, t_2) \int_0^\infty \int_0^\infty A_g(t_1, u) A_{3-g}(t_2, v) \\ & \times G_{g(3-g)}\{(t_1, u), (t_2, v)\}\, \mathrm{d}v\, \mathrm{d}u\end{aligned}$$

becomes the asymptotic covariance for $\mathscr{T}(t_1)$ and $\mathscr{T}(t_2)$. Taking $t_1 = t_2 = t$ provides $\sigma^2(t)$, the variance of $\mathscr{T}(t)$ in the case of a single analysis as found in Murray [4].

A2. Estimation of variances and covariances in this paper

All asymptotic closed form variance and covariance terms in this paper are easily estimated. However, additional notation is required. First note that in estimating joint and conditional quantities in relation to group g_1 at time t_1 and group g_2 at time t_2, attention is restricted to those $k = 1, \ldots, n_{g_1g_2}(t_1, t_2)$ correlated pairs where study entry has occurred for both members of the pair at their respective analysis times. In estimating marginal quantities in relation to group g at time t, all individual pair members entered into group g before time t will be considered regardless of whether their correlated counterpart has been entered into the study.

Let $Y_{g_1g_2}\{(t_1, x_1), (t_2, x_2)\} = \sum_{k=1}^{n_{g_1g_2}(t_1,t_2)} I(X_{g_1k}(t_1) \geqslant x_1, X_{g_2k}(t_2) \geqslant x_2)$ count the number of correlated pairs where the pair member in group g_1 at analysis time t_1 is still at risk at study time x_1 and the pair member in group g_2 at analysis time t_2 is still at risk at study time x_2. Also, let $\mathrm{d}N_{g_1g_2}\{(t_1, x_1), (t_2, x_2)\} = \sum_{k=1}^{n_{g_1g_2}(t_1,t_2)} I(x_1 \leqslant X_{g_1k}(t_1) < x_1 + \Delta x_1, x_2 \leqslant X_{g_2k}(t_2) < x_2 + \Delta x_2, \Delta_{g_1k}(t_1) = 1, \Delta_{g_2k}(t_2) = 1)$ count the number of correlated pairs where the pair member in group g_1 at analysis time t_1 fails at study time x_1 and the pair member in group g_2 at analysis time t_2 fails at study time x_2. Let $\mathrm{d}N_{g_1|g_2}\{(t_1, x_1)|(t_2, x_2)\} = \sum_{k=1}^{n_{g_1g_2}(t_1,t_2)} I(x_1 \leqslant X_{g_1k}(t_1) < x_1 + \Delta x_1, X_{g_2k}(t_2) \geqslant x_2, \Delta_{g_1k}(t_1) = 1)$ count the number of correlated pairs where the pair member in group g_1 at analysis time t_1 had been at risk until failing at study time x_1 and the pair member in group g_2 at analysis time t_2 remains at risk at study time x_2. An unpooled estimate for $G_{g_1g_2}\{(t_1, x_1), (t_2, x_2)\}\, \mathrm{d}x_1\, \mathrm{d}x_2$ becomes

$$\hat{G}_{g_1g_2}\{(t_1, x_1), (t_2, x_2)\} = \frac{n_{g_1}(t_1) n_{g_2}(t_2)}{n_{g_1g_2}(t_1, t_2)} \frac{Y_{g_1g_2}\{(t_1, x_1), (t_2, x_2)\}}{Y_{g_1}(t_1, x_1) Y_{g_2}(t_2, x_2)} \left[\frac{\mathrm{d}N_{g_1g_2}\{(t_1, x_1), (t_2, x_2)\}}{Y_{g_1g_2}\{(t_1, x_1), (t_2, x_2)\}}\right.$$

$$-\frac{\mathrm{d}N_{g_1|g_2}\{(t_1,x_1)|(t_2,x_2)\}\mathrm{d}N_{g_2}(t_2,x_2)}{Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}Y_{g_2}(t_2,x_2)}$$

$$-\frac{\mathrm{d}N_{g_2|g_1}\{(t_2,x_2)|(t_1,x_1)\}\mathrm{d}N_{g_1}(t_2,x_1)}{Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}Y_{g_1}(t_2,x_1)}+\frac{\mathrm{d}N_{g_1}(t_2,x_1)\mathrm{d}N_{g_2}(t_2,x_2)}{Y_{g_1}(t_2,x_1)Y_{g_2}(t_2,x_2)}\Bigg]$$

An unpooled estimate for $A_g(t,x)$ is $\hat{A}_g(t,x)=\int_x^\infty J(t,u)\hat{S}_g(t,u)\,\mathrm{d}u$. Hence an unpooled estimate for $\sigma^2(t)$ becomes

$$\hat{\sigma}^2(t)=\sum_{g=1}^{2}\frac{\hat{\pi}_1(t)\hat{\pi}_2(t)}{\hat{\pi}_g(t)}\left[\int_0^\infty n_g(t)\frac{\{\hat{A}_g(t,u)\}^2\,\mathrm{d}N_g(t,u)}{\{Y_g(t,u)\}^2}\right]$$

$$-\hat{\theta}(t)\int_0^\infty\int_0^\infty\hat{A}_1(t,u)\hat{A}_2(t,v)\hat{G}_{12}\{(t,u),(t,v)\}$$

Under the null hypothesis, pooling may be employed in estimating elements of $\sigma^2(t)$. Pooling time-to-event data available from those entered prior to time t from groups g_1 and g_2, define $\tilde{S}(t,x)$ and $K\tilde{M}(t,x)$ as the pooled right-continuous and left-continuous versions of the Kaplan–Meier estimator of the survivor function, respectively, relating to study time x. Define $\hat{H}_g(t,x)$ as the Kaplan–Meier estimate of the left-continuous version of the censoring survival function for group g at analysis time t. Define $\tilde{A}(t,x)=\int_0^\infty J(t,u)\tilde{S}(t,u)\,\mathrm{d}u$. 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 $G_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}\,\mathrm{d}x_1\,\mathrm{d}x_2$ becomes

$$\tilde{G}_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}=\frac{Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}}{n_{g_1g_2}(t_1,t_2)K\tilde{M}(t_1,x_1)K\tilde{M}(t_2,x_2)\hat{H}_{g_1}(t_1,x_1)\hat{H}_{g_2}(t_2,x_2)}$$

$$\times\left[\frac{\mathrm{d}N_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}}{Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}}-\frac{\mathrm{d}N_{g_1|g_2}\{(t_1,x_1)|(t_2,x_2)\}\mathrm{d}\bar{N}(t_2,x_2)}{Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}\bar{Y}(t_2,x_2)}\right.$$

$$\left.-\frac{\mathrm{d}N_{g_2|g_1}\{(t_2,x_2)|(t_1,x_1)\}\mathrm{d}\bar{N}(t_2,x_1)}{Y_{g_1g_2}\{(t_1,x_1),(t_2,x_2)\}\bar{Y}(t_2,x_1)}+\frac{\mathrm{d}\bar{N}(t_2,x_1)\,\mathrm{d}\bar{N}(t_2,x_2)}{\bar{Y}(t_2,x_1)\bar{Y}(t_2,x_2)}\right]$$

Hence a pooled estimate for $\sigma^2(t)$ becomes

$$\tilde{\sigma}^2(t)=\sum_{g=1}^{2}\frac{\hat{\pi}_1(t)\hat{\pi}_2(t)}{\hat{\pi}_g(t)}\left[\int_0^\infty\frac{\{\tilde{A}(t,u)\}^2\,\mathrm{d}\bar{N}(t,u)}{\hat{H}_g(t,u)K\tilde{M}(t,u)\bar{Y}(t,u)}\right]$$

$$-\hat{\theta}(t)\int_0^\infty\int_0^\infty\tilde{A}(t,u)\tilde{A}(t,v)\tilde{G}_{12}\{(t,u),(t,v)\}$$

For use in the covariance term, $\sigma(t_1,t_2)$, define $\hat{A}_g(t_1,t_2,x)=\int_x^\infty J(t_1,u)\hat{S}_g(t_2,u)\,\mathrm{d}u$ and $\tilde{A}_g(t_1,t_2,x)=\int_x^\infty J(t_1,u)\tilde{S}_g(t_2,u)\,\mathrm{d}u$ so that the most updated information available is used in

estimating $S(x)$ within these terms. An unpooled estimate for $\sigma(t_1, t_2)$ becomes

$$\hat{\sigma}(t_1,t_2) = \sum_{g=1}^{2} \{\hat{\pi}_{3-g}(t_1)\hat{\pi}_{3-g}(t_2)\hat{\pi}_g(t_1|t_2)\}^{\frac{1}{2}} \int_0^\infty n_g(t_2) \frac{\hat{A}_g(t_1,t_2,u)\hat{A}_g(t_2,t_2,u)\,\mathrm{d}N_g(t_2,u)}{\{Y_g(t_2,u)\}^2}$$
$$-\sum_{g=1}^{2} \hat{\psi}_{g(3-g)}(t_1,t_2) \int_0^\infty \int_0^\infty \hat{A}_g(t_1,t_2,u)\hat{A}_{3-g}(t_2,t_2,v)\hat{G}_{g(3-g)}\{(t_1,u),(t_2,v)\}$$

An estimate that pools data under the null hypothesis would be

$$\tilde{\sigma}(t_1,t_2) = \sum_{g=1}^{2} \{\hat{\pi}_{3-g}(t_1)\hat{\pi}_{3-g}(t_2)\hat{\pi}_g(t_1|t_2)\}^{\frac{1}{2}} \int_0^\infty \frac{\tilde{A}(t_1,t_2,u)\tilde{A}(t_2,t_2,u)\,\mathrm{d}\bar{N}(t_2,u)}{\hat{H}_g(t_2,u)K\tilde{M}(t_2,u)\bar{Y}(t_2,u)}$$
$$-\sum_{g=1}^{2} \hat{\psi}_{g(3-g)}(t_1,t_2) \int_0^\infty \int_0^\infty \tilde{A}(t_1,t_2,u)\tilde{A}(t_2,t_2,v)\tilde{G}_{g(3-g)}\{(t_1,u),(t_2,v)\}$$

ACKNOWLEDGEMENTS

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

REFERENCES

1. Murray S, Tsiatis AA. Sequential methods for comparing years of life saved in the two-sample censored data problem. *Biometrics* 1999; **55**:1085–1092.
2. Li Z. A group sequential test for survival trials: an alternative to rank-based procedures. *Biometrics* 1999; **55**:277–283.
3. Pepe MS, Fleming TR. Weighted Kaplan–Meier statistics – a class of distance tests for censored survival data. *Biometrics* 1989; **45**:497–507.
4. Murray S. Using Weighted Kaplan–Meier statistics in nonparametric comparisons of paired censored survival outcomes. *Biometrics* 2001; **57**:361–368.
5. Early Treatment Diabetic Retinopathy Study Research Group. Early Treatment Diabetic Retinopathy Study design and baseline patient characteristics: ETDRS report number 7. *Ophthalmology* 1991; **98**:741–756.
6. Early Treatment Diabetic Retinopathy Study Research Group. Early Photocoagulation for Diabetic Retinopathy: ETDRS report number 9. *Ophthalmology* 1991; **98**:766–785.
7. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977; **64**:191–199.
8. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979; **35**:549–556.
9. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika* 1983; **70**:659–663.
10. Chang I, Hsiung C, Chuang Y. Applications of a frailty model to sequential survival analysis. *Statistica Sinica* 1997; **7**:127–138.
11. Murray S. Nonparametric rank-based methods for group sequential monitoring of paired censored survival data. *Biometrics* 2000; **56**:984–990.
12. Lin DY. Nonparametric sequential testing in clinical trials with incomplete multivariate observations. *Biometrika* 1991; **78**:123–131.
13. Muñoz SR, Bangdiwala SI, Sen PK. Group sequential methods for censored bivariate survival data. *Brazilian Journal of Probability and Statistics* 1997; **11**:11–25.

