Web-based Supplementary Materials for Nonparametric Group Sequential Methods for Evaluating Survival Benefit from Multiple Short-Term Follow-up Windows

Meng Xia[a], Susan Murray[a], Nabihah Tayob[b]
[a] *University of Michigan, Department of Biostatistics, Ann Arbor, MI 48109, USA*
[b] *The University of Texas MD Anderson Cancer Center, Department of Biostatistics,Houston, TX 77030, USA*

1 Web Appendix A: Derivation of the asymptotic joint distribution of the proposed test statistic at interim analysis times, $s_1, \ldots, s_K$

In this section, we derive the asymptotic joint distribution of the proposed test statistics, $\mathscr{T}(s_1), \ldots, \mathscr{T}(s_K)$ at interim analysis times, $s_1, \ldots, s_K$. The overall strategy is to first show that the vector of test statistics, $\{\mathscr{T}(s_1), \ldots, \mathscr{T}(s_K)\}$, is asymptotically equivalent in distribution to the more tractable vector of random variables, $\{\mathscr{T}^*(s_1), \ldots, \mathscr{T}^*(s_K)\}$, where elements of this latter vector are based on sums of independent and identically distributed quantities. From there, a standard application of the multivariate central limit theorem gives the desired result.

Our test statistic at analysis time s,

$$\mathscr{T}(s) = \sqrt{\frac{n_1(s)n_2(s)}{n_1(s) + n_2(s)}}\{\hat{\mu}_1(s, \tau) - \hat{\mu}_2(s, \tau)\},$$

can be rewritten as

$$\mathscr{T}(s) = \sqrt{\frac{n_2(s)}{n_1(s)+n_2(s)}}\sqrt{n_1(s)}\hat{\mu}_1(s,\tau) - \sqrt{\frac{n_1(s)}{n_1(s)+n_2(s)}}\sqrt{n_2(s)}\hat{\mu}_2(s,\tau), \tag{1}$$

where $n_g(s)/\{n_1(s)+n_2(s)\} \xrightarrow{p} \pi_g(s)$. Suppose at analysis time s, combining information across b follow-up windows of length τ, we record M events $\{0 \equiv T_0 < T_1 < ... < T_M < T_{M+1} \equiv \tau\}$. Then, by Taylor series expansion,

$$\sqrt{n_g(s)}\hat{\mu}_g(s,\tau) = \sqrt{n_g(s)}\sum_{m=0}^{M}(T_{m+1}-T_m)exp\left\{-\sum_{j=0}^{m}\frac{dN_g(s,T_j)}{Y_g(s,T_j)}\right\}$$

is asymptotically equivalent in distribution to the following terms:

$$\sqrt{n_g(s)}\sum_{m=0}^{M}(T_{m+1}-T_m)\,exp\left\{-\sum_{j=0}^{m}\lambda_g^W(s,T_j)dT_j\right\} \tag{2}$$

$$+\sqrt{n_g(s)}\sum_{m=0}^{M}(T_{m+1}-T_m)\left[\sum_{j=0}^{m}-exp\left\{-\sum_{j'=0}^{m}\lambda_g^W(s,T_{j'})dT_{j'}\right\}\frac{dN_g(s,T_j)}{Y_g(s,T_j)}\right] \tag{3}$$

$$-\sqrt{n_g(s)}\sum_{m=0}^{M}(T_{m+1}-T_m)\left[\sum_{j=0}^{m}-exp\left\{-\sum_{j'=0}^{m}\lambda_g^W(s,T_{j'})dT_{j'}\right\}\lambda_g^W(s,T_j)dT_j\right] \tag{4}$$

$$+\sqrt{n_g(s)}\sum_{m=0}^{M}(T_{m+1}-T_m)\frac{1}{2!}exp\left\{-\sum_{j'=0}^{m}\lambda_g^W(s,T_{j'})dT_{j'}\right\}\left[\sum_{j=0}^{m}\left\{\frac{dN_g(s,T_j)}{Y_g(s,T_j)}-\lambda_g^W(s,T_j)dT_j\right\}\right]^2 \tag{5}$$

$$+\sqrt{n_g(s)}\sum_{m=0}^{M}(T_{m+1}-T_m)[\text{higher order terms}] \tag{6}$$

Using arguments similar to those in Tayob and Murray (2016) Appendix B, terms (5) and (6) converge to zero in probability. When there is no treatment effect (i.e., the null hypothesis is true), terms (2) and (4) for group $g = 1$ will cancel the corresponding terms for group $g = 2$ in the test statistic $\mathscr{T}(s)$. Hence, the asymptotic distribution of $\mathscr{T}(s)$ is based on the behavior of term (3) for groups $g = 1, 2$. Term (3) can be further rewritten as

$$\sqrt{n_g(s)} \sum_{m=0}^{M} (T_{m+1} - T_m) \left[\sum_{j=0}^{m} -exp \left\{ - \sum_{j'=0}^{m} \lambda_g^W(s, T_{j'}) dT_{j'} \right\} \frac{dN_g(s, T_j)}{Y_g(s, T_j)} \right]$$

$$= -\sqrt{n_g(s)} \sum_{m=0}^{M} (T_{m+1} - T_m) exp\{- \sum_{j'=0}^{m} \lambda_g^W(s, T_{j'}) dT_{j'}\} \sum_{j=0}^{m} \frac{dN_g(s, T_j)}{Y_g(s, T_j)},$$

which is asymptotically equivalent in distribution (via Taylor series) to

$$-\sqrt{n_g(s)} \sum_{m=0}^{M} (T_{m+1} - T_m) exp \left\{ - \sum_{j'=0}^{m} \lambda_g^W(s, T_{j'}) dT_{j'} \right\} \times$$

$$\left\{ \sum_{j=0}^{m} \frac{EdN_g(s, T_j)}{EY_g(s, T_j)} \right. \tag{7}$$

$$+ \sum_{j=0}^{m} \left[\frac{1}{EY_g(s, T_j)} [dN_g(s, T_j) - EdN_g(s, T_j)] - \frac{EdN_g(s, T_j)}{EY_g(s, T_j)^2} [Y_g(s, T_j) - EY_g(s, T_j)] \right] \tag{8}$$

$$\left. + [\text{higher order terms}] \right\}. \tag{9}$$

Using arguments similar to those in Tayob and Murray Appendix B once again, the higher order terms in (9) converge to zero in probability. In addition when the null hypothesis is true, term (7) for group $g = 1$ will cancel with its counterpart term for $g = 2$ in the test statistic $\mathscr{T}(s)$. Hence, the asymptotic distribution of $\mathscr{T}(s)$ is based on the behavior of term (8) for groups $g = 1, 2$ which upon noting that $EdN_g(s, T_j)/EY_g(s, T_j) = \lambda_g^W(s, T_j)$ and $EY_g(s, T_j) = \sum_{l=1}^{b} Pr(X_{gi}(s, t_l) \geq T_j)$ can be algebraically rearranged as:

$$-\sqrt{n_g(s)} \sum_{m=0}^{M} (T_{m+1} - T_m) exp \left\{ - \sum_{j'=0}^{m} \lambda_g^W(s, T_{j'}) dT_{j'} \right\} \sum_{j=0}^{m} \frac{dN_g(s, T_j) - Y_g(s, T_j) \lambda_g^W(s, T_j)}{\sum_{l=1}^{b} Pr(X_{gi}(s, t_l) \geq T_j)}$$

or returning to more standard stochastic integral notation as:

$$-\sqrt{n_g(s)}\int_0^{\tau} exp\left\{-\int_0^{u_2}\lambda_g^W(s,u_1)du_1\right\}\int_0^{u_2}\frac{dN_g(s,u_1)-Y_g(s,u_1)\lambda_g^W(s,u_1)}{\sum_{l=1}^b Pr(X_{gi}(s,t_l)\geq u_1)}du_2. \quad (10)$$

Summarizing calculations from equation (1) to equation (10),

$$\mathscr{T}(s)=\sqrt{\frac{n_2(s)}{n_1(s)+n_2(s)}}\sqrt{n_1(s)}\hat{\mu}_1(s,\tau)-\sqrt{\frac{n_1(s)}{n_1(s)+n_2(s)}}\sqrt{n_2(s)}\hat{\mu}_2(s,\tau)$$

is asymptotically equivalent in distribution to

$$\sqrt{\pi_1(s)}\sqrt{n_2(s)}\int_0^{\tau} exp\left\{-\int_0^{u_2}\lambda_2^W(s,u_1)du_1\right\}\int_0^{u_2}\frac{dN_2(s,u_1)-Y_2(s,u_1)\lambda_2^W(s,u_1)}{\sum_{l=1}^b Pr(X_{2i}(s,t_l)\geq u_1)}du_2$$

$$-\sqrt{\pi_2(s)}\sqrt{n_1(s)}\int_0^{\tau} exp\left\{-\int_0^{u_2}\lambda_1^W(s,u_1)du_1\right\}\int_0^{u_2}\frac{dN_1(s,u_1)-Y_1(s,u_1)\lambda_1^W(s,u_1)}{\sum_{l=1}^b Pr(X_{1i}(s,t_l)\geq u_1)}du_2.$$

From here, we note that the remaining terms above can be written in terms of independent and identically distributed random variables that lend themselves to standard limiting distribution results via the multivariate central limit theorem. Recall that

$$N_g(s,u)=\sum_{i=1}^{n_g(s)}N_{gi}(s,u)=\sum_{i=1}^{n_g(s)}\sum_{j=1}^{b}N_{gi}(s,t_j,u)$$

and

$$Y_g(s,u)=\sum_{i=1}^{n_g(s)}Y_{gi}(s,u)=\sum_{i=1}^{n_g(s)}\sum_{j=1}^{b}Y_{gi}(s,t_j,u).$$

Define:

$$Z_{ij}\{\hat{\mu}_g(s,\tau)\}=\int_0^{\tau} exp\left\{-\int_0^{u_2}\lambda_g^W(s,u_1)du_1\right\}\int_0^{u_2}\frac{dN_{gi}(s,t_j,u_1)-Y_{gi}(s,t_j,u_1)\lambda_g^W(s,u_1)du_1}{\sum_{l=1}^b Pr(X_{gi}(s,t_l)\geq u_1)}du_2$$

and

$$Z_i\{\hat{\mu}_g(s,\tau)\}=\sum_{j=1}^{b}Z_{ij}\{\hat{\mu}_g(s,\tau)\}.$$

Note that $Z_i\{\hat{\mu}_g(s,\tau)\}$ only depends on patient i and is independent and identically distributed for $i = 1, \ldots, n_g(s)$. Using this notation, the above asymptotically equivalent representation of the distribution of $\mathscr{T}(s)$ becomes

$$\mathscr{T}^*(s) = \sqrt{\pi_1(s)}\sqrt{n_2(s)}\frac{\sum_{i=1}^{n_2(s)} Z_i\{\hat{\mu}_2(s,\tau)\}}{n_2(s)} - \sqrt{\pi_2(s)}\sqrt{n_1(s)}\frac{\sum_{i=1}^{n_1(s)} Z_i\{\hat{\mu}_1(s,\tau)\}}{n_1(s)}. \quad (11)$$

Application of the multivariate central limit theorem to the vector of test statistics $\{\mathscr{T}^*(s_1), \ldots, \mathscr{T}^*(s_K)\}$ calculated at calendar times, $s_1, s_2, \ldots, s_K$ (K finite), gives a limiting multivariate normal distribution as $n_g(s_1) \to \infty, g = 1, 2$, with asymptotic covariance matrix estimated empirically as described in Web Appendix B. A closed-form version of the asymptotic covariance is described in Web Appendix C.

For convenience, we explicitly describe the special case where only a single analysis is performed. When the null hypothesis is true, the asymptotic limiting distribution of $\mathscr{T}(s)$ is Normal with mean 0 and variance $\pi_2(s)\sigma_1^2(s) + \pi_1(s)\sigma_2^2(s)$, where $\sigma_g^2(s), g = 1, 2$ is the variance of $Z_i(\hat{\mu}_g(s,\tau))$ and can be estimated using the sampling variability of $Z_i\{\hat{\mu}_g(s,\tau)\}$, that is, $\hat{\sigma}_g^2(s) = \sum_{i=1}^{n_g(s)}[z_i\{\hat{\mu}_g(s,\tau)\} - \bar{z}\{\hat{\mu}_g(s,\tau)\}]^2/[n_g(s) - 1]$, where

$$z_i\{\hat{\mu}_g(s,\tau)\} = \sum_{j=1}^{b} z_{ij}\{\hat{\mu}_g(s,\tau)\}; \qquad \bar{z}\{\hat{\mu}_g(s,\tau)\} = \sum_{i=1}^{n_g(s)} z_i\{\hat{\mu}_g(s,\tau)\}/n_g(s)$$

and

$$z_{ij}\{\hat{\mu}_g(s,\tau)\} = \int_0^\tau exp\left\{-\int_0^{u_2} \frac{dN_g(s,u_1)}{Y_g(s,u_1)}\right\}\left\{\int_0^{u_2} \frac{dN_{gi}(s,t_j,u_1) - Y_{gi}(s,t_j,u_1)\frac{dN_g(s,u_1)}{Y_g(s,u_1)}}{Y_g(s,u_1)/n_g(s)}\right\} du_2. (12)$$

For finite sample sizes, we use a standardized version of the test statistic,

$$\tilde{\mathscr{T}}(s) = \frac{\mathscr{T}(s)}{\sqrt{\hat{\pi}_2(s)\hat{\sigma}_1^2(s) + \hat{\pi}_1(s)\hat{\sigma}_2^2(s)}},$$

which has an approximate Normal(0,1) distribution, with critical values of $\pm$ 1.96 conferring an overall type I error of 5% when a single analysis is performed.

2 Web Appendix B: Empirical covariance matrix for $\left\{\tilde{\mathscr{T}}(s_1), \ldots, \tilde{\mathscr{T}}(s_K)\right\}$

In this appendix, we describe how to estimate the empirical version of the $K \times K$ asymptotic covariance matrix, Σ, corresponding to standardized test statistics, $\left\{\tilde{\mathscr{T}}(s_1), \ldots, \tilde{\mathscr{T}}(s_K)\right\}$. By design, diagonal elements of this matrix are equal to one, so that this covariance matrix is also a correlation matrix. Off-diagonal elements, $\sigma_{k_1k_2} = \sigma_{k_2k_1}, k_1 < k_2$, can be estimated based on the more updated dataset at analysis s_{k_2}.

In Web Appendix A, we show that $\{\mathscr{T}(s_1), \ldots, \mathscr{T}(s_K)\}$ is asymptotically equivalent in distribution to $\{\mathscr{T}^*(s_1), \ldots, \mathscr{T}^*(s_K)\}$. Similarly, for the standardized version of each test statistic, $\tilde{\mathscr{T}}(s_k), s_k = s_1, \ldots, s_K$, we work with the corresponding asymptotically equivalent in distribution standardized form, $\mathscr{T}^*(s_k)/\sqrt{\pi_2(s_k)\sigma_1^2(s_k) + \pi_1(s_k)\sigma_2^2(s_k)}$. Hence, off-diagonal elements $\sigma_{k_1k_2} = \sigma_{k_2k_1}, k_1 < k_2$, of the covariance matrix, Σ, can be estimated by

$$\hat{\sigma}_{k_1k_2} = \frac{\hat{Cov}\left\{\mathscr{T}^*(s_{k_1}), \mathscr{T}^*(s_{k_2})\right\}}{\sqrt{\hat{\pi}_2(s_{k_1})\tilde{\sigma}_1^2(s_{k_1}) + \hat{\pi}_1(s_{k_1})\tilde{\sigma}_2^2(s_{k_1})}\sqrt{\hat{\pi}_2(s_{k_2})\hat{\sigma}_1^2(s_{k_2}) + \hat{\pi}_1(s_{k_2})\hat{\sigma}_2^2(s_{k_2})}} \tag{13}$$

We define each component of $\hat{\sigma}_{k_1k_2}$ in more detail below.

Estimated terms that use the most up-to-date information at analysis time s_{k_2} have already been described for $\hat{\sigma}_g^2(s_{k_2}), g = 1, 2$, in Web Appendix A, captured by terms in equation (12). Web Appendix A also defines $\hat{\pi}_g(s_k) = n_g(s_k)/\{n_1(s_k) + n_2(s_k)\}$, for $g = 1, 2$ and $s_k = s_{k_1}, s_{k_2}$. Estimates of $\sigma_g^2(s_{k_1}), g = 1, 2$ used in the covariance estimate are modified to take advantage of additional information available at s_{k_2} for estimating terms that do not depend on analysis time. In particular, since both $dN_{gi}(s_{k_1}, t_j, u_1)/Y_{gi}(s_{k_1}, t_j, u_1)$ and $dN_{gi}(s_{k_2}, t_j, u_1)/Y_{gi}(s_{k_2}, t_j, u_1)$ estimate $\lambda_{gi}(t_j, u_1)du_1$, and the latter term uses more data, we replace $dN_{gi}(s_{k_1}, t_j, u_1)$ in equation (1) with $Y_{gi}(s_{k_1}, t_j, u_1) \times\ dN_{gi}(s_{k_2}, t_j, u_1)/Y_{gi}(s_{k_2}, t_j, u_1)$. Similarly in equation (12), we replace $Y_g(s_{k_1}, u_1)/n_g(s_{k_1})$, which is an estimate of $\sum_{l=1}^{b} Pr(T_{gi}(s_{k_1}, t_l) \geq u_1) Pr(C_{gi}(s_{k_1}, t_l) \geq u_1)$, with $\left[\sum_{i=1}^{n_g(s_{k_2})} I\{T_{gi} \geq u_1 + t_l\}/n_g(s_{k_2})\right]$ $\times \left[\sum_{i=1}^{n_g(s_{k_1})} I\{C_{gi}(s_{k_1}) \geq u_1 + t_l\}/n_g(s_{k_1})\right]$. Here, terms involving the event time are estimated using updated data, while terms involving the censoring distribution remain relevant to analysis time s_{k_1}. Putting these modifications together gives us

$$\tilde{z}_{ij}\{\hat{\mu}_g(s_{k_1}, \tau)\} = \int_0^{\tau} exp\{-\int_0^{u_2} \frac{dN_g(s_{k_1}, u_1)}{Y_g(s_{k_1}, u_1)}\} \Bigg[\int_0^{u_2} \left\{ \sum_{l=1}^{b} \left(\sum_{i=1}^{n_g(s_{k_2})} I\{T_{gi} \geq u_1 + t_l\} \sum_{i\prime=1}^{n_g(s_{k_1})} I\{C_{gi\prime}(s_{k_1}) \geq u_1 + t_l\} \right) \right\}^{-1} \times n_g(s_{k_1}) n_g(s_{k_2}) Y_{gi}(s_{k_1}, t_j, u_1) \left\{ \frac{dN_{gi}(s_{k_2}, t_j, u_1)}{Y_{gi}(s_{k_2}, t_j, u_1)} - \frac{dN_g(s_{k_1}, u_1)}{Y_g(s_{k_1}, u_1)} \right\} \Bigg] du_2$$

as an updated version of $z_{ij}\{\hat{\mu}_g(s_{k_1}, \tau)\}$ for use in covariance terms. And mimicking Web Appendix A notation, $\tilde{\sigma}_g^2(s_{k_1})$ used in equation (13) is calculated by replacing $z_{ij}\{\hat{\mu}_g(s_{k_1}, \tau)\}$ with $\tilde{z}_{ij}\{\hat{\mu}_g(s_{k_1}, \tau)\}$ terms in corresponding formulas for $\hat{\sigma}_g^2(s_{k_1})$ from Web Appendix A.

The only remaining undefined term from equation (13) is $\hat{Cov}\left\{\mathscr{T}^*(s_{k_1}), \mathscr{T}^*(s_{k_2})\right\}$, which is described in the following. From equation (11),

$$Cov\left\{\mathscr{T}^*(s_{k_1}), \mathscr{T}^*(s_{k_2})\right\}$$

$$= \sum_{g=1}^{2} Cov\left[\sqrt{\pi_{3-g}(s_{k_1})}\sqrt{n_g(s_{k_1})}\frac{\sum_{i=1}^{n_g(s_{k_1})} Z_i\{\hat{\mu}_g(s_{k_1}, \tau)\}}{n_g(s_{k_1})}, \sqrt{\pi_{3-g}(s_{k_2})}\sqrt{n_g(s_{k_2})}\frac{\sum_{i=1}^{n_g(s_{k_2})} Z_i\{\hat{\mu}_g(s_{k_2}, \tau)\}}{n_g(s_{k_2})}\right].$$

Without loss of generality, assume $k_1 \leq k_2$ so that $n_g(s_{k_1}) \leq n_g(s_{k_2})$ and there are $n_g(s_{k_1})$ patients contributing (correlated) data from both analysis times. Then the previous expression reduces to

$$= \sum_{g=1}^{2} \sqrt{\pi_{3-g}(s_{k_1})}\sqrt{\pi_{3-g}(s_{k_2})}\frac{n_g(s_{k_1})}{\sqrt{n_g(s_{k_1}) n_g(s_{k_2})}} Cov\left[Z_i\{\hat{\mu}_g(s_{k_1}, \tau)\}, Z_i\{\hat{\mu}_g(s_{k_2}, \tau)\}\right],$$

which is asymptotically equivalent to

$$= \sum_{g=1}^{2} \sqrt{\pi_{3-g}(s_{k_1})\pi_{3-g}(s_{k_2})\psi_g(s_{k_1}, s_{k_2})} Cov\left[Z_i\{\hat{\mu}_g(s_{k_1}, \tau)\}, Z_i\{\hat{\mu}_g(s_{k_2}, \tau)\}\right]$$

where $\psi_g(s_{k_1}, s_{k_2})$ is the limiting proportion of patients entered at s_{k_1} of those eventually entered by s_{k_2} of group g, that is estimated by $n_g(s_{k_1})/n_g(s_{k_2})$.

In practice, $Cov\left[Z_i\{\hat{\mu}_g(s_{k_1}, \tau)\}, Z_i\{\hat{\mu}_g(s_{k_2}, \tau)\}\right]$ can be estimated based on the empirical covariance of sample realizations of $Z_i\{\hat{\mu}_g(s_{k_1}, \tau)\}$ and $Z_i\{\hat{\mu}_g(s_{k_2}, \tau)\}$, that is,

$$\hat{Cov}\left[Z_i\{\hat{\mu}_g(s_{k_1}, \tau)\}, Z_i\{\hat{\mu}_g(s_{k_2}, \tau)\}\right] = \sum_{i=1}^{n_g(s_{k_1})} \frac{[\tilde{z}_i\{\hat{\mu}_g(s_{k_1}, \tau)\} - \bar{\tilde{z}}\{\hat{\mu}_g(s_{k_1}, \tau)\}][z_i\{\hat{\mu}_g(s_{k_2}, \tau)\} - \bar{z}\{\hat{\mu}_g(s_{k_2}, \tau)\}]}{n_g(s_{k_1}) - 1},$$

where $\tilde{z}_i\{\hat{\mu}_g(s_{k_1}, \tau)\} = \sum_{j=1}^{b} \tilde{z}_{ij}\{\hat{\mu}_g(s_{k_1}, \tau)\}$, $\bar{\tilde{z}}\{\hat{\mu}_g(s_{k_1}, \tau)\} = \sum_{i=1}^{n_g(s_{k_1})} \tilde{z}_i\{\hat{\mu}_g(s_{k_1}, \tau)\}/n_g(s_{k_1})$. Putting each described component into equation (13), we have the version of $\hat{\sigma}_{k_1 k_2}$ listed in Section 4 of the main manuscript.

3 Web Appendix C: Closed form covariance matrix for $\left\{\tilde{\mathscr{T}}(s_1), \ldots, \tilde{\mathscr{T}}(s_K)\right\}$

At times it is convenient to have an asymptotic closed form version of the covariance matrix for $\left\{\tilde{\mathscr{T}}(s_1), \ldots, \tilde{\mathscr{T}}(s_K)\right\}$, for instance in assessing whether an independent increments variability structure is present. Working from results in the last paragraph of Web Appendix B, instead of estimating $Cov\left[Z_i\{\hat{\mu}_g(s_{k_1}, \tau)\}, Z_i\{\hat{\mu}_g(s_{k_2}, \tau)\}\right]$ with the empirical covariance, in this section we derive its asymptotic closed-form formula. Consider $Z_i\{\hat{\mu}_g(s_k, \tau)\} = \sum_{j=1}^{b} Z_{ij}\{\hat{\mu}_g(s_k, \tau)\}$ at analysis times $s_k = s_{k_1}$ and s_{k_2} and recall that group g patients are independent and identically distributed. Then

$$Cov\left[Z_i\{\hat{\mu}_g(s_{k_1}, \tau)\}, Z_i\{\hat{\mu}_g(s_{k_2}, \tau)\}\right] = \sum_{j=1}^{b} \sum_{j\prime=1}^{b} Cov\left[Z_{ij}\{\hat{\mu}_g(s_{k_1}, \tau)\}, Z_{ij\prime}\{\hat{\mu}_g(s_{k_2}, \tau)\}\right].$$

For notational simplicity, we submerge the group indicator g as we work with the summand term above. That is,

$$\begin{aligned}
&Cov\left[Z_{ij}\{\hat{\mu}_g(s_{k_1},\tau)\}, Z_{ij\prime}\{\hat{\mu}_g(s_{k_2},\tau)\}\right] \\
&= \sum_{j=1}^{b}\sum_{j\prime=1}^{b}\int_0^\tau\int_0^\tau exp\{-\int_0^{u_2}\lambda^W(s_{k_1},u_1)du_1\}exp\{-\int_0^{v_2}\lambda^W(s_{k_2},v_1)dv_1\} \\
&\quad\times\int_0^{u_2}\int_0^{v_2}\frac{1}{\sum_l Pr(X_i(s_{k_1},t_l)\geq u_1)\sum_{l\prime}Pr(X_i(s_{k_2},t_{l\prime})\geq v_1)} \\
&\quad\times Cov\Big\{dN_i(s_{k_1},t_j,u_1)-Y_i(s_{k_1},t_j,u_1)\lambda^W(s_{k_1},u_1)du_1, dN_i(s_{k_2},t_{j\prime},v_1) \\
&\qquad\qquad -Y_i(s_{k_2},t_{j\prime},v_1)\lambda^W(s_{k_2},v_1)dv_1\Big\}du_2dv_2.
\end{aligned}$$

Focusing on this last term:

$$Cov\Big\{dN_i(s_{k_1},t_j,u_1)-Y_i(s_{k_1},t_j,u_1)\lambda^W(s_{k_1},u_1)du_1, dN_i(s_{k_2},t_{j\prime},v_1)-Y_i(s_{k_2},t_{j\prime},v_1)\lambda^W(s_{k_2},v_1)dv_1\Big\}$$

$$= E[dN_i(s_{k_1},t_j,u_1)dN_i(s_{k_2},t_{j\prime},v_1)] \tag{14}$$

$$-\lambda^W(s_{k_1},u_1)E[Y_i(s_{k_1},t_j,u_1)dN_i(s_{k_2},t_{j\prime},v_1)]du_1 \tag{15}$$

$$-\lambda^W(s_{k_2},v_1)E[Y_i(s_{k_2},t_{j\prime},v_1)dN_i(s_{k_1},t_j,u_1)]dv_1 \tag{16}$$

$$+\lambda^W(s_{k_1},u_1)\lambda^W(s_{k_2},v_1)E[Y_i(s_{k_1},t_j,u_1)Y_i(s_{k_2},t_{j\prime},v_1)]du_1dv_1 \tag{17}$$

$$-E[dN_i(s_{k_1},t_j,u_1)-Y_i(s_{k_1},t_j,u_1)\lambda^W(s_{k_1},u_1)du_1] \tag{18}$$

$$\times E[dN_i(s_{k_2},t_{j\prime},v_1)-Y_i(s_{k_2},t_{j\prime},v_1)\lambda^W(s_{k_2},v_1)dv_1]. \tag{19}$$

Term (14) becomes:

$$E[dN_i(s_{k_1},t_j,u_1)dN_i(s_{k_2},t_{j'},v_1)] = \lim_{\Delta u_1,\Delta v_1\to 0} Pr\{u_1 \le X_i(s_{k_1},t_j) < u_1+\Delta u_1, \delta_i(s_{k_1},t_j)=1,$$
$$v_1 \le X_i(s_{k_2},t_{j'}) < v_1+\Delta v_1, \delta_i(s_{k_2},t_{j'})=1\}$$
$$= \lim_{\Delta u_1\to 0} Pr\{u_1 \le X_i(s_{k_1},t_j) < u_1+\Delta u_1, \delta_i(s_{k_1},t_j)=1\}$$
$$\times I\{u_1+t_j = v_1+t_{j'}\}$$
$$= \lambda(s_{k_1},t_j,u_1)Pr\{X_i(s_{k_1},t_j)\ge u_1\}I\{u_1+t_j=v_1+t_{j'}\}du_1.$$

Term (15) becomes

$$E[Y_i(s_{k_1},t_j,u_1)dN_i(s_{k_2},t_{j'},v_1)] = \lim_{\Delta v_1\to 0} Pr\{X_i(s_{k_1},t_j)\ge u_1, v_1 \le X_i(s_{k_2},t_{j'}) < v_1+\Delta v_1,$$
$$\delta_i(s_{k_2},t_{j'})=1\}$$
$$= \lambda(s_{k_2},t_{j'},v_1)Pr\{X_i(s_{k_1},t_j)\ge u_1, X_i(s_{k_2},t_{j'})\ge v_1\}$$
$$[I\{u_1+t_j \le v_1+t_{j'}\} + I\{u_1=0, t_j > v_1+t_{j'}\}]dv_1,$$

where the expectation is only none-zero when $u_1+t_j \le v_1+t_{j'}$. The term $I\{u_1=0, t_j > v_1+t_{j'}\}$ comes from the case when the failure occurs before calendar time t_j, namely $t_j > v_1+t_{j'}$, by definition $X_i(s_{k_1},t_j)=0$. Therefore the expectation is also non-zero when $u_1=0$.

Term (16) becomes

$$E[Y_i(s_{k_2},t_{j'},v_1)dN_i(s_{k_1},t_j,u_1)] = \lim_{\Delta u_1\to 0} Pr\{X_i(s_{k_2},t_{j'})\ge v_1, u_1 \le X_i(s_{k_1},t_j) < u_1+\Delta u_1,$$
$$\delta_i(s_{k_1},t_j)=1\}$$
$$= \lambda(s_{k_1},t_j,u_1)Pr\{X_i(s_{k_1},t_j)\ge u_1, X_i(s_{k_2},t_{j'})\ge v_1\}$$
$$[I\{u_1+t_j \ge v_1+t_{j'}\} + I\{v_1=0, u_1+t_j < t_{j'}\}]du_1,$$

where the expectation is only none-zero when $u_1+t_j \geq v_1+t_{j'}$. The term $I\{v_1=0, u_1+t_j < t_{j'}\}$ comes from the case when the failure occurs before calendar time $t_{j'}$, namely $u_1+t_j < t_{j'}$, by definition $X_i(s_{k_2}, t_{j'}) = 0$. Therefore the expectation is also non-zero when $v_1 = 0$.

Term (17) becomes

$$E[Y_i(s_{k_1}, t_j, u_1)Y_i(s_{k_2}, t_{j'}, v_1)] = Pr\{X_i(s_{k_1}, t_j) \geq u_1, X_i(s_{k_2}, t_{j'}) \geq v_1\}.$$

Term (18) becomes

$$\begin{aligned} E[dN_i(s_{k_1}, t_j, u_1) - Y_i(s_{k_1}, t_j, u_1)\lambda^W(s_{k_1}, u_1)du_1] =& [\lambda(s_{k_1}, t_j, u_1) - \lambda^W(s_{k_1}, u_1)] \\ & \times Pr\{X_i(s_{k_1}, t_j) \geq u_1\}du_1. \end{aligned}$$

And term (19) becomes

$$\begin{aligned} E[dN_i(s_{k_2}, t_{j'}, v_1) - Y_i(s_{k_2}, t_{j'}, v_1)\lambda^W(s_{k_2}, v_1)dv_1] =& [\lambda(s_{k_2}, t_{j'}, v_1) - \lambda^W(s_{k_2}, v_1)] \\ & \times Pr\{X_i(s_{k_2}, t_{j'}) \geq v_1\}dv_1. \end{aligned}$$

Substituting appropriate terms we now have

$$
\begin{aligned}
&Cov\,[Z_i\{\hat{\mu}_g(s_{k_1},\tau)\}, Z_i\{\hat{\mu}_g(s_{k_2},\tau)\}] \\
&= \sum_{j=1}^{b}\sum_{j\prime=1}^{b}\int_0^{\tau}\int_0^{\tau} exp\{-\int_0^{u_2}\lambda_g^W(s_{k_1},u_1)du_1\}exp\{-\int_0^{v_2}\lambda_g^W(s_{k_2},v_1)dv_1\} \\
&\quad\times \int_0^{u_2}\int_0^{v_2}\frac{1}{\sum_l Pr(X_{gi}(s_{k_1},t_l)\geq u_1)\sum_{l\prime} Pr(X_{gi}(s_{k_2},t_{l\prime})\geq v_1)} \\
&\quad\times \Bigg\{\lambda_g(t_j,u_1)Pr\{X_{gi}(s_{k_1},t_j)\geq u_1\}I\{u_1+t_j=v_1+t_{j\prime}\}du_1 \\
&\quad - \Bigg[\lambda_g^W(s_{k_1},u_1)\lambda(t_{j\prime},v_1)[I\{u_1+t_j\leq v_1+t_{j\prime}\}+I\{u_1=0,t_j>v_1+t_{j\prime}\}] \\
&\qquad + \lambda_g^W(s_{k_2},v_1)\lambda(t_j,u_1)[I\{u_1+t_j\geq v_1+t_{j\prime}\}+I\{v_1=0,u_1+t_j<t_{j\prime}\}] \\
&\qquad - \lambda_g^W(s_{k_1},u_1)\lambda^W(s_{k_2},v_1)\Bigg]Pr\{X_{gi}(s_{k_1},t_j)\geq u_1, X_{gi}(s_{k_2},t_{j\prime})\geq v_1\}du_1dv_1 \\
&\quad - \{\lambda_g(t_j,u_1)-\lambda_g^W(s_{k_1},u_1)\}\{\lambda_g(t_{j\prime},v_1)-\lambda_g^W(s_{k_2},v_1)\} \\
&\qquad \times Pr\{X_{gi}(s_{k_1},t_j)\geq u_1\}Pr\{X_{gi}(s_{k_2},t_{j\prime})\geq v_1\}du_1dv_1\Bigg\}du_2dv_2
\end{aligned}
$$

Unfortunately, this covariance does not simplify to an independent increments structure except in special cases such as an exponentially distributed event time. The independent increments structure emerges in this special case upon noting that $\lambda_g^W(s,u)=\lambda(t,u)=\lambda$ for all s,t and u. However, given the advantages of avoiding parametric assumptions, there is no practical computation savings that can be made from knowledge of this special case.

We've also used this asymptotic closed form variance as a method to double-check that R code for our empirically calculated covariance is on target. For example, a covariance matrix estimated from 500 individuals' data should be relatively close to the asymptotic closed form. Assuming an $Exp(0.5)$ event time with 2 years of uniform accrual, and analyses using $\tau=1$ conducted at 1, 2, 3, 4 and 5 years in calendar time, the closed form covariance matrix calculation gives:

$$\begin{bmatrix} 0.175 & 0.085 & 0.042 & 0.039 & 0.035 \\ 0.085 & 0.091 & 0.052 & 0.041 & 0.036 \\ 0.042 & 0.052 & 0.054 & 0.042 & 0.038 \\ 0.039 & 0.041 & 0.042 & 0.043 & 0.038 \\ 0.035 & 0.036 & 0.038 & 0.038 & 0.039 \end{bmatrix},$$

whereas the corresponding empirical covariance estimate from the 500 individuals was

$$\begin{bmatrix} 0.149 & 0.082 & 0.050 & 0.040 & 0.034 \\ 0.082 & 0.104 & 0.056 & 0.044 & 0.039 \\ 0.050 & 0.056 & 0.059 & 0.045 & 0.041 \\ 0.040 & 0.044 & 0.045 & 0.046 & 0.042 \\ 0.034 & 0.039 & 0.041 & 0.042 & 0.043 \end{bmatrix}$$

with difference matrix

$$\begin{bmatrix} -0.026 & -0.003 & 0.008 & 0.001 & -0.001 \\ -0.003 & 0.013 & 0.004 & 0.003 & 0.003 \\ 0.008 & 0.004 & 0.005 & 0.003 & 0.003 \\ 0.001 & 0.003 & 0.003 & 0.003 & 0.004 \\ -0.001 & 0.003 & 0.003 & 0.004 & 0.004 \end{bmatrix}.$$

Repeating this exercise for different simulated datasets and sample sizes is a comforting coding check.

4 Web Appendix D: Supplemental simulation results

In this section we show supplemental simulation results for our proposed method using the same simulation scenarios 1-9 described in the main manuscript. In Web Tables S1 and S2, we (1) examine the performance of our method for alternative choices of $\tau = 0.25, 0.50$ and 0.75 years, (2) show results for the Peto and Peto (WLR-PP) test that places more weight on hazards at the beginning of the study and (3) show results for the Fleming-Harrington (WLR-FH) $(0.5, 0.5)$ test that places more weight on hazards at the end of the study. Web Table S1 shows stopping rates based on OF efficacy, JT safety, Pocock safety and OF safety bounds. Web Table S2 shows the average study time (AST) in years, the average sample number (ASN) and the average number of events (ANE).

All test statistic boundaries meet their targets within simulation error under Scenario 1, the null hypothesis (Web Table S1, Scenario 1).

For the most part, stopping rates do not seem to vary much based on the selection for τ. The only possible exception is in Scenario 4, the delayed treatment effect scenario, where power is slightly smaller for smaller values of τ. The WLR-FH test does well in this setting, with slightly less power than the proposed test using $\tau = 1$ year and slightly more power than the proposed test with smaller values of τ. The WLR-PP test has much lower power than all other methods in this setting. The WLR-PP test also performs poorly in Scenario 8, the Scenario with mixed cure distribution alternatives under consideration.

Note that these extra simulations for $\tau = 0.25, 0.5$ and 0.75 are not intended to be an exhaustive look at how to choose τ since we believe most applications will have a natural choice. But these additional simulations verify that the method performs well for a broader selection of short-term window lengths.

5 Web Appendix E: Supplemental example results

Web Figure S1 shows group sequential OF efficacy boundaries as well as OF, Pocock and JT safety boundaries for the proposed test statistic (left panel), the RMS statistic (middle panel) and the logrank statistic (right panel). All test statistics are standardized to ease comparisons between panels of the figure. Boundaries and test statistics shown in Web Figure S1 are enumerated for clarity in Web Table S3. Although historically during that period of clinical trial design symmetric stopping boundaries were typically used, a more modern safety boundary would make sense in this setting, particularly since it was not known for certain that the low-dose was sufficient to protect against mortality in the same way the high dose had up to that time. Observed values of the test statistics in each panel of Web Figure S1 are superimposed as dots with bold connecting lines. None of the test statistics approached the safety boundaries at any of the interim analyses. As shown in Web Table S3, the standardized proposed test statistics and the standardized RMS test statistics crossed the OF efficacy boundary at year 1990. The logrank test did not cross the OF efficacy boundary at any interim analysis time.

Table S1: Rates of stopping for efficacy or for safety

Scenario	Test Statistic	OF Efficacy	JT Safety	P Safety	OF Safety
1	Proposed $\tau = 0.75$	0.024	0.192	0.025	0.024
	Proposed $\tau = 0.5$	0.023	0.197	0.024	0.023
	Proposed $\tau = 0.25$	0.024	0.193	0.024	0.024
	WLR-PP	0.023	0.195	0.025	0.026
	WLR-FH $(0.5, 0.5)$	0.023	0.196	0.026	0.026
2	Proposed $\tau = 0.75$	0.813	0	0	0
	Proposed $\tau = 0.5$	0.803-0.804	0.002	0	0
	Proposed $\tau = 0.25$	0.806-0.807	0.002	0	0
	WLR-PP	0.75	0	0	0
	WLR-FH $(0.5, 0.5)$	0.807	0	0	0
3	Proposed $\tau = 0.75$	0	0.977	0.79	0.847
	Proposed $\tau = 0.5$	0	0.979	0.773	0.829
	Proposed $\tau = 0.25$	0	0.973	0.778	0.839
	WLR-PP	0	0.967	0.724	0.76
	WLR-FH $(0.5, 0.5)$	0	0.971	0.773	0.815
4	Proposed $\tau = 0.75$	0.813-0.825	0.024	0.007	0
	Proposed $\tau = 0.5$	0.817-0.824	0.019	0.005	0
	Proposed $\tau = 0.25$	0.803-0.811	0.025	0.007	0
	WLR-PP	0.367	0.029	0.008	0
	WLR-FH $(0.5, 0.5)$	0.823-0.834	0.031	0.008	0
5	Proposed $\tau = 0.75$	0	0.970	0.742	0.817
	Proposed $\tau = 0.5$	0	0.967	0.730	0.819
	Proposed $\tau = 0.25$	0	0.965	0.718	0.809
	WLR-PP	0	0.743	0.325	0.381
	WLR-FH $(0.5, 0.5)$	0	0.970	0.778	0.848
6	Proposed $\tau = 0.75$	0.764	0	0	0
	Proposed $\tau = 0.5$	0.767	0.001	0	0
	Proposed $\tau = 0.25$	0.761	0	0	0
	WLR-PP	0.753	0.001	0	0
	WLR-FH $(0.5, 0.5)$	0.784	0	0	0
7	Proposed $\tau = 0.75$	0	0.960	0.707	0.744
	Proposed $\tau = 0.5$	0	0.961	0.701	0.748
	Proposed $\tau = 0.25$	0	0.959	0.696	0.743
	WLR-PP	0	0.956	0.696	0.736
	WLR-FH $(0.5, 0.5)$	0	0.963	0.725	0.768
8	Proposed $\tau = 0.75$	0.883	0	0	0
	Proposed $\tau = 0.5$	0.876	0	0	0
	Proposed $\tau = 0.25$	0.879	0	0	0
	WLR-PP	0.777	0	0	0
	WLR-FH $(0.5, 0.5)$	0.854	0	0	0
9	Proposed $\tau = 0.75$	0	0.991	0.849	0.890
	Proposed $\tau = 0.5$	0	0.989	0.840	0.881
	Proposed $\tau = 0.25$	0	0.988	0.843	0.884
	WLR-PP	0	0.959	0.735	0.774
	WLR-FH $(0.5, 0.5)$	0	0.982	0.807	0.852

Table S2: AST in years, ASN and ANE in Scenarios 1 - 9

		AST			**ASN**			**ANE**		
Scenario	**Test Statistic**	**JT**	**P**	**OF**	**JT**	**P**	**OF**	**JT**	**P**	**OF**
1	Proposed $\tau = 0.75$	4.7	4.9	5.0	195	199	200	156	163	164
	Proposed $\tau = 0.5$	4.7	4.9	5.0	195	199	200	156	163	165
	Proposed $\tau = 0.25$	4.7	4.9	5.0	195	199	200	156	164	165
	WLR-PP	4.7	4.9	5.0	195	199	200	156	163	165
	WLR-FH $(0.5, 0.5)$	4.7	4.9	5.0	195	199	200	156	163	164
2	Proposed $\tau = 0.75$	3.8	3.8	3.8	186	186	186	144	144	144
	Proposed $\tau = 0.5$	3.8	3.8	3.8	185	185	185	144	144	144
	Proposed $\tau = 0.25$	3.8	3.8	3.8	186	186	186	145	145	145
	WLR-PP	3.8	3.8	3.8	186	186	186	145	145	145
	WLR-FH $(0.5, 0.5)$	3.7	3.7	3.7	184	184	184	142	142	142
3	Proposed $\tau = 0.75$	2.1	3.0	3.7	151	169	185	93	120	142
	Proposed $\tau = 0.5$	2.1	3.1	3.7	151	170	185	93	123	143
	Proposed $\tau = 0.25$	2.1	3.1	3.7	152	170	185	94	123	144
	WLR-PP	2.1	3.1	3.8	152	170	185	94	123	144
	WLR-FH $(0.5, 0.5)$	2.1	3.0	3.6	152	169	184	94	121	141
4	Proposed $\tau = 0.75$	3.9	3.9	4.0	189	190	190	135	137	138
	Proposed $\tau = 0.5$	3.9	3.9	4.0	189	190	190	136	137	138
	Proposed $\tau = 0.25$	3.9	4.0	4.0	190	191	191	137	139	140
	WLR-PP	4.5	4.6	4.7	195	196	197	152	154	155
	WLR-FH $(0.5, 0.5)$	3.7	3.8	3.8	186	188	188	132	134	135
5	Proposed $\tau = 0.75$	2.9	3.9	4.0	171	186	191	110	135	140
	Proposed $\tau = 0.5$	3.0	3.9	4.0	171	187	191	111	135	140
	Proposed $\tau = 0.25$	3.0	3.9	4.1	172	187	192	112	136	141
	WLR-PP	3.7	4.5	4.6	182	194	197	129	151	154
	WLR-FH $(0.5, 0.5)$	2.8	3.7	3.9	168	184	189	107	132	137
6	Proposed $\tau = 0.75$	3.7	3.7	3.7	184	184	184	143	143	143
	Proposed $\tau = 0.5$	3.7	3.7	3.7	184	184	184	144	144	144
	Proposed $\tau = 0.25$	3.7	3.7	3.7	185	185	185	145	145	145
	WLR-PP	3.7	3.7	3.7	184	184	184	145	145	145
	WLR-FH $(0.5, 0.5)$	3.6	3.6	3.6	183	183	183	142	142	142
7	Proposed $\tau = 0.75$	2.2	3.1	3.7	152	170	184	95	124	144
	Proposed $\tau = 0.5$	2.2	3.2	3.8	153	170	185	96	125	145
	Proposed $\tau = 0.25$	2.2	3.2	3.8	152	171	185	95	126	146
	WLR-PP	2.2	3.2	3.8	152	171	185	95	126	145
	WLR-FH $(0.5, 0.5)$	2.1	3.1	3.7	152	170	184	94	124	144
8	Proposed $\tau = 0.75$	3.5	3.5	3.5	181	181	181	129	129	129
	Proposed $\tau = 0.5$	3.5	3.5	3.5	181	181	181	129	129	129
	Proposed $\tau = 0.25$	3.5	3.5	3.5	182	182	182	130	130	130
	WLR-PP	3.7	3.7	3.7	184	184	184	133	133	133
	WLR-FH $(0.5, 0.5)$	3.5	3.5	3.5	181	181	181	129	129	129
9	Proposed $\tau = 0.75$	2.1	3.0	3.5	151	169	183	91	114	131
	Proposed $\tau = 0.5$	2.1	3.0	3.6	152	170	183	91	114	131
	Proposed $\tau = 0.25$	2.1	3.0	3.5	151	170	183	91	115	131
	WLR-PP	2.2	3.2	3.8	153	171	185	92	116	134
	WLR-FH $(0.5, 0.5)$	2.1	3.0	3.6	152	170	183	91	115	131

Table S3: Test statistics and efficacy or safety boundaries

	Proposed				**RMS**				**Logrank**			
	1987	1988	1989	1990	1987	1988	1989	1990	1987	1988	1989	1990
Test Statistics	0.96	1.62	2.20	2.12	0.92	1.97	2.27	2.20	0.92	1.76	2.00	1.83
OF Efficacy	3.91	2.78	2.31	2.00	3.92	2.78	2.28	1.97	3.91	2.77	2.31	1.99
JT Safety	-1.96	-1.65	-1.36	-1.02	-1.96	-1.61	-1.26	-0.95	-1.96	-1.64	-1.35	-1.01
Pocock Safety	-2.37	-2.46	-2.44	-2.30	-2.37	-2.41	-2.33	-2.17	-2.37	-2.44	-2.43	-2.28
OF Safety	-3.91	-2.78	-2.31	-2.00	-3.92	-2.78	-2.28	-1.97	-3.91	-2.77	-2.31	-1.99

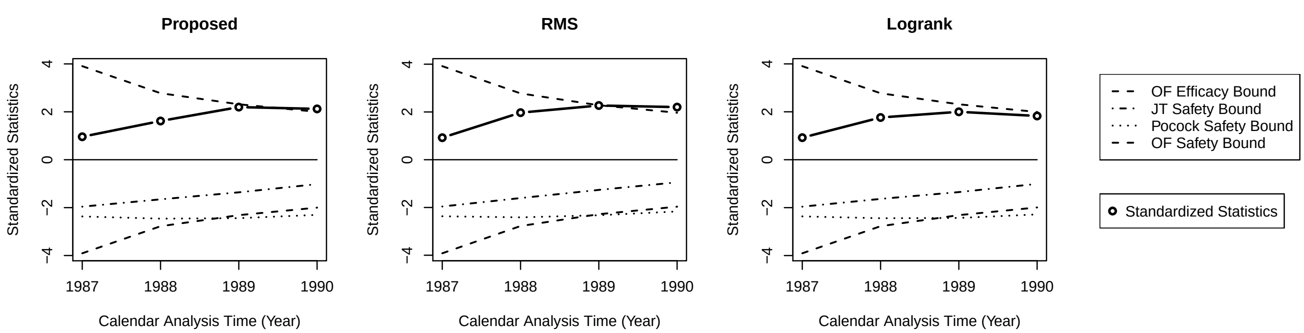

Figure S1: Standardized test statistics and stopping boundaries
(RMS: Restricted Mean Survival; OF:O'Brien and Fleming; JT: Jennison and Turnbull)