1 Methods

1.1 Weighted log-rank statistics

Suppose there are n subjects in the the study, each of which has an event time T_i and a censoring time C_i. We assume that the event time has a survival function S_k with density f_k and hazard function λ_k and the censoring time follows a survival function H_k, provided that the subject was assigned to group k (i.e. the group indicator $Z_i = k$), where $k = 1$ denotes the treatment group and $k = 0$ the control group. As usual, we assume that T_i and C_i are independent given Z_i, $i = 1, \dots, n$. Statistical inference will be based on the observed data $\{Y_i = \min(T_i, C_i), \delta_i = I(Y_i = T_i), Z_i, i = 1, \dots, n\}$.

We often want to test the null hypothesis $H_0(\tau) : \lambda_1(s) = \lambda_0(s)$ for all $s \le \tau$, which is equivalent to test $S_1(s) = S_0(s)$ for all $s \le \tau$.

To test the null hypothesis, weighted log-rank statistic is often used, which takes the form

$$\mathcal{L}_W(\tau) = n^{-1} \sum_{i=1}^{n} \int_0^\tau W(s) \Big\{ Z_i - \frac{\Gamma_1(s)}{\Gamma(s)} \Big\} dN_i(s), \qquad (1)$$

where, for any s, $\Gamma_k(s) = n^{-1} \sum_{i=1}^n I(Y_i \ge s, Z_i = k)$, $k = 0, 1$, $\Gamma(s) = \Gamma_1(s) + \Gamma_0(s)$, $N_i(s) = I(T_i \le s, \delta_i = 1)$, $i = 1, \dots, n$. Different types of weighted log-rank statistics can be specified based on the choice of the possibly data-dependent weight function $W(s)$.

LR. If $W(s) = 1$, $\mathcal{L}_W(\tau)$ is the (un-weighted) log-rank test statistic.

Gehan. If $W(s) = \Gamma(s)$, $\mathcal{L}_W(\tau)$ is the Gehan test statistic.

TW. If $W(s) = \Gamma^{1/2}(s)$, $\mathcal{L}_W(\tau)$ is the Tarone-Ware test statistic.

FH. Let $\widehat{S}$ be the Kaplan-Meier estimator of the survival function in the pooled sample, i.e.

$$\widehat{S}(t) = \prod_{s \le t} \Big\{ 1 - \frac{dN(s)}{\Gamma(s)} \Big\},$$

where $N(s) = n^{-1} \sum_{i=1}^n N_i(s)$. If $W(s) = [\widehat{S}(s-)]^\rho [1 - \widehat{S}(s-)]^\gamma$ (Note, $\widehat{S}(t-)$ is $\widehat{S}$ evaluated at the time point right before t.) for $\rho \ge 0, \gamma \ge 0$, then $\mathcal{L}_W(\tau)$ is the Fleming-Harrington FH(ρ, γ) test statistic (Fleming and Harrington, 1981).

PP. A slightly modified version of the Kaplan-Meier estimator $\widetilde{S}$ is defined as

$$\widetilde{S}(t) = \prod_{s \le t} \Big\{ 1 - \frac{dN(s)}{\Gamma(s) + 1/n} \Big\}.$$

If $W(s) = \widetilde{S}(s)$, then $\mathcal{L}_W(\tau)$ is the Peto-Peto test statistic, and a modified version of Peto-Peto test statistic is with $W(s) = \widetilde{S}(s)\Gamma(s)/\{\Gamma(s) + 1/n\}$.

BFY. To handle rare event, Buyske, Fagerstrom and Ying (2000) proposed to modify the Fleming-Harrington $G^{\rho,0}$ test statistic by using a weight $W(s) = [\widehat{S}(s-) - \widehat{S}(\tau-)]^\rho$.

YP. Yang and Prentice (2005) proposed a two-parameter model for the hazard ratio

$$\frac{\lambda_1(s)}{\lambda_0(s)} = \frac{\theta_1 \theta_2}{\theta_1 + (\theta_2 - \theta_1) S_0(s)}, \qquad s \le \tau,$$

where θ_1 and θ_2 are the so-called short-term and long-term hazard ratios respectively. With this model, Yang and Prentice (2010) further proposed to use weighted functions W_1 and W_2, where $W_1 = 1/W_2$ and $W_2(s)$ is the estimated hazard ratios with θ_1,θ_2 and $S_0(t)$ being consistently estimated. Even though Yang and Prentice (2010) showed that the resulting weighted log-rank tests are valid (asymptotically), some simulation shows that it may have inflated type-1 error (Chauvel and O'Quigley, 2014).

Under $H_0(\tau)$ and suitable regularity conditions, $\sqrt{n}\mathcal{L}_W(\tau)$ converges weakly to a normal distribution with mean zero and variance $V_W(\tau)$. The variance can be consistently estimated by

$$\mathcal{V}_W(\tau) = \int_0^\tau W^2(s)\frac{\Gamma_1(s)\Gamma_0(s)}{\Gamma^2(s)}dN(s).$$

The weighted log-rank test can be generalized to perform parameter estimation for the overall average effect size. In this regard, we may write the score function as

$$\mathcal{L}_W(\beta,\tau) = n^{-1}\sum_{i=1}^n \int_0^\tau W(s)\Big\{Z_i - \frac{\Gamma_1(\beta,s)}{\Gamma(\beta,s)}\Big\}dN_i(s), \tag{2}$$

where, for any s, $\Gamma_k(\beta,s) = n^{-1}\sum_{i=1}^n \exp\{\beta Z_i\}I(Y_i \geq s, Z_i = k)$, $k = 0,1$, $\Gamma(\beta,s) = \Gamma_1(\beta,s) + \Gamma_0(\beta,s)$. Apparently, $\mathcal{L}_W(0,\tau) = \mathcal{L}_W(\tau)$. Let $\widehat{\beta}_W(\tau)$ be the solution of $\mathcal{L}_W(\beta,\tau) = 0$. Under suitable regularity conditions, $\sqrt{n}\{\widehat{\beta}_W(\tau) - \beta_W(\tau)\}$ converges weakly to a normal distribution with mean zero and a sandwich-form variance $U_W^{-1}\{\beta_W(\tau),\tau\}V_W\{\beta_W(\tau),\tau\}U_W^{-1}\{\beta_W(\tau),\tau\}$. Here $\beta_W(\tau)$ is the solution of $L_W(\beta,\tau) = 0$, where

$$L_W(\beta,\tau) = n^{-1}\sum_{i=1}^n \int_0^\tau w(s)\Big[E\{Z_i dN_i(s)\} - \frac{E\{\Gamma_1(\beta,s)\}}{E\{\Gamma(\beta,s)\}}dE\{N_i(s)\}\Big],$$

$w(s)$ is the limit of $W(s)$. For any fixed β, the variance $U_W(\beta,\tau)$ can be consistently estimated by

$$\mathcal{U}_W(\beta,\tau) = \int_0^\tau W(s)\frac{\Gamma_1(\beta,s)\Gamma_0(\beta,s)}{\Gamma^2(\beta,s)}dN(s),$$

and the variance $V_W(\beta,\tau) = \text{var}\{\sqrt{n}\mathcal{L}_W(\beta,\tau)\}$ can be consistently estimated by the Cox-model-based estimator (Sasieni, 1993a)

$$\mathcal{U}_{W^2}(\beta,\tau) = \int_0^\tau W^2(s)\frac{\Gamma_1(\beta,s)\Gamma_0(\beta,s)}{\Gamma^2(\beta,s)}dN(s).$$

Alternatively, $V_W(\beta,\tau)$ can be estimated using a robust approach (Lin and Wei, 1989; Sasieni, 1993b) as

$$\mathcal{V}_W(\beta,\tau) = n^{-1}\sum_{i=1}^n \Big\{\mathcal{A}_{1i}(\beta,\tau) - \mathcal{A}_{2i}(\beta,\tau)\Big\}^2,$$

where

$$\mathcal{A}_{1i}(\beta,\tau) = \int_0^\tau W(s)\left[Z_i - \frac{\Gamma_1(\beta,s)}{\Gamma(\beta,s)}\right]dN_i(s), \tag{3}$$

$$\mathcal{A}_{2i}(\beta,\tau) = \int_0^\tau W(s)\left[Z_i - \frac{\Gamma_1(\beta,s)}{\Gamma(\beta,s)}\right]I(Y_i \geq s)\exp\{\beta Z_i\}\frac{dN(s)}{\Gamma(\beta,s)}. \tag{4}$$

Recently, in order to quantify the non-constant hazard ratio across time, Lin and Leon (2017) proposed to use the following model for testing and estimation

$$\lambda_1(s) = \exp\{\beta\Phi(s)\}\lambda_0(s), s \leq \tau, \tag{5}$$

where $\Phi(s)$ is a known or pre-specified function such that it is bounded in $[0, \tau]$ and the maximum point corresponds to where the maximum treatment effect can be obtained and anywhere else denotes the places with reduced treatment effects. This is essentially a time-dependent Cox model with the time-dependent covariate at time s equal to $Z\Phi(s)$. Based on the model, a testing and estimating procedure can be developed. Specifically, let $\widetilde{\beta}_\Phi(\tau)$ be the solution of $\mathcal{Q}_\Phi(\beta, \tau) = 0$, where

$$\mathcal{Q}_\Phi(\beta, \tau) = n^{-1}\sum_{i=1}^{n}\int_0^\tau \Big\{Z_i\Phi(s) - \frac{\widetilde{\Gamma}_1(\Phi, \beta, s)}{\widetilde{\Gamma}_0(\Phi, \beta, s)}\Big\}dN_i(s), \tag{6}$$

and, for any Φ, β and s, $\widetilde{\Gamma}_q(\Phi, \beta, s) = n^{-1}\sum_{i=1}^{n}\{Z_i\Phi(s)\}^q \exp\{\beta\Phi(s)Z_i\}I(Y_i \geq s)$, $q = 0, 1, 2$. Lin and Leon (2017) showed that

$$\mathcal{Q}_\Phi(0, \tau) = n^{-1}\sum_{i=1}^{n}\int_0^\tau \Phi(s)\Big\{Z_i - \frac{\Gamma_1(s)}{\Gamma(s)}\Big\}dN_i(s),$$

which is a weighted log-rank statistic with weight $W(s) = \Phi(s)$. Note that this weight is not data-dependent. Under suitable regularity conditions, $\sqrt{n}\{\widetilde{\beta}_\Phi(\tau) - \beta_\Phi(\tau)\}$ converges weakly to a normal distribution with mean zero and a sandwich-form variance

$$\widetilde{U}_\Phi^{-1}\{\beta_\Phi(\tau), \tau\}\widetilde{V}_\Phi\{\beta_\Phi(\tau), \tau\}\widetilde{U}_\Phi^{-1}\{\beta_\Phi(\tau), \tau\}.$$

Here $\beta_\Phi(\tau)$ is the solution of $Q_\Phi(\beta, \tau) = 0$, where

$$Q_\Phi(\beta, \tau) = n^{-1}\sum_{i=1}^{n}\int_0^\tau \Big[E\{Z_i\Phi(s)dN_i(s)\} - \frac{E\{\widetilde{\Gamma}_1(\Phi, \beta, s)\}}{E\{\widetilde{\Gamma}_0(\Phi, \beta, s)\}}dE\{N_i(s)\}\Big].$$

For any fixed β, the variance $\widetilde{U}_\Phi(\beta, \tau)$ can be consistently estimated by

$$\widetilde{\mathcal{U}}_\Phi(\beta, \tau) = \int_0^\tau \Big[\frac{\widetilde{\Gamma}_2(\Phi, \beta, s)}{\widetilde{\Gamma}_0(\Phi, \beta, s)} - \frac{\widetilde{\Gamma}_1^2(\Phi, \beta, s)}{\widetilde{\Gamma}_0^2(\Phi, \beta, s)}\Big]dN(s)$$

and the variance $\widetilde{V}_\Phi(\beta, \tau) = \text{var}\{\sqrt{n}\mathcal{Q}_\Phi(\beta, \tau)\}$, which, based on the model (5), can be consistently estimated by $\widetilde{\mathcal{U}}_\Phi(\beta, \tau)$. Alternatively, a robust estimator of $\widetilde{V}_\Phi(\beta, \tau)$ is

$$\widetilde{\mathcal{V}}_\Phi(\beta, \tau) = n^{-1}\sum_{i=1}^{n}\Big\{\mathcal{B}_{1i}(\beta, \tau) - \mathcal{B}_{2i}(\beta, \tau)\Big\}^2,$$

where

$$\mathcal{B}_{1i}(\beta, \tau) = \int_0^\tau \left[Z_i\Phi(s) - \frac{\widetilde{\Gamma}_1(\Phi, \beta, s)}{\widetilde{\Gamma}_0(\Phi, \beta, s)}\right]dN_i(s), \tag{7}$$

$$\mathcal{B}_{2i}(\beta, \tau) = \int_0^\tau \left[Z_i\Phi(s) - \frac{\widetilde{\Gamma}_1(\Phi, \beta, s)}{\widetilde{\Gamma}_0(\Phi, \beta, s)}\right]I(Y_i \geq s)\exp\{\beta\Phi(s)Z_i\}\frac{dN(s)}{\widetilde{\Gamma}_0(\Phi, \beta, s)}. \tag{8}$$

We have a few remarks on the above methods.

R1. The differences $\mathcal{A}_{1i}\{\widehat{\beta}_W(\tau),\tau\} - \mathcal{A}_{2i}\{\widehat{\beta}_W(\tau),\tau\}$ and $\mathcal{B}_{1i}\{\widetilde{\beta}_\Phi(\tau),\tau\} - \mathcal{B}_{2i}\{\widetilde{\beta}_\Phi(\tau),\tau\}$, $i = 1,\ldots,n$, are the score residuals in the corresponding models, which are in the standard outputs of many existing software. We can therefore use them to provide an easy estimation of the asymptotic variances.

R2. The cut-point τ can be ∞, which is the case for the log-rank, Gehan and Tarone-Ware statistics. We use a cut-point τ in the general case to avoid a lengthy discussion of the tails. In the general case, a usual requirement for τ is to have $E\{\Gamma(\tau)\} > 0$.

1.2 Kaplan-Meier based statistics

Let $\widehat{S}_k$ be the Kaplan-Meier estimator of the survival function in group $k = 0, 1$. Another class of non-parametric tests, called the weighted Kaplan-Meier tests (Pepe and Fleming, 1989, 1991), can be expressed as

$$\mathcal{K}_\Psi(\tau) = \int_0^\tau \{\widehat{S}_1(s) - \widehat{S}_0(s)\} d\Psi(s), \tag{9}$$

where Ψ is a known but possibly data-dependent function of bounded variation in $[0,\tau]$. If $\Psi(t) = t$, then the test statistic is the difference of the restricted mean survival times (RMSTs). If $\Psi(t) = I(t \geq \tau)$ then the test statistic reduces to $\widehat{S}_1(\tau) - \widehat{S}_0(\tau)$. If $\Psi(t) = \int_0^t W(s)ds$, the test statistic is the weighted difference of Kaplan-Meier curves. Pepe and Fleming (1989) used the weight

$$W(s) = \frac{\widehat{H}_1(s-)\widehat{H}_0(s-)}{p_1\widehat{H}_1(s-) + p_0\widehat{H}_0(s-)}$$

where $\widehat{H}_k$ is the Kaplan-Meier estimate of the survival function H_k for the censoring time, $p_k = n_k/n$ and n_k is the sample size in group k, $k = 0, 1$. This test is based on a linear combination of weighted differences of the two Kaplan-Meier curves over time and is a natural tool to assess the difference of the two survival functions directly by using an inverse probability censoring function to improve stability in the tails. Uno et al. (2015) proposed to use a data-adaptive weight function $W(s)$ that is proportional to the difference $\widehat{S}_1(s) - \widehat{S}_0(s)$. Because the weight is data-driven, the improvement in power over Pepe and Flemings weight function is universal and robust.

Let $\Psi_0(t)$ be the limit of $\Psi(t)$ and

$$K_{\Psi_0}(\tau) = \int_0^\tau \{S_1(s) - S_0(s)\} d\Psi_0(s),$$

then under $H_0(\tau)$ and some regularity conditions, $\sqrt{n}\{\mathcal{K}_\Psi(\tau) - K_{\Psi_0}(\tau)\}$ converges in distribution to a mean-zero normal distribution with the variance $\sigma_1^2(\tau) + \sigma_0^2(\tau)$. For $k = 0, 1$, $\sigma_k^2(\tau)$ can be consistently estimated by

$$\widehat{\sigma}_k^2(\tau) = \int_0^\tau \left\{ \int_s^\tau \widehat{S}_k(u) d\Psi(u) \right\}^2 \frac{dN_k(s)}{\Gamma_k(s)\{\Gamma_k(s) - \Delta N_k(s)\}},$$

where $N_k(s) = n^{-1}\sum_{i=1}^n I(T_i \leq s, \delta_i = 1, Z_i = k)$.

The RMST has attracted a lot of interests in recent years due to its simple and intuitive interpretation as the treatment effect measure. This effect measure does not reply on the assumptions such as the proportional hazards assumption. Specifically, let $\mathcal{R}_k(\tau) = \int_0^\tau \widehat{S}_k(s)ds$ be the estimator of the RMST

$R_k(\tau) = E\{\text{min}(T,\tau) \mid Z = k\} = \int_0^\tau S_k(s)ds$ in group $k = 0, 1$. Similar to the Kaplan-Meier curves, one can visually examine the difference between the two RMST curves $\mathcal{R}_k(\tau)$ at each cut-points τ (Zhao et al. 2016). If any particular cut-point τ is of interest, one may consider the difference of the RMSTs $\mathcal{RD}(\tau) = \mathcal{R}_1(\tau) - \mathcal{R}_0(\tau)$, the ratio of the RMSTs $\mathcal{RR}(\tau) = \mathcal{R}_1(\tau)/\mathcal{R}_0(\tau)$, and the ratio of time losses $\mathcal{RL}(\tau) = \{\tau - \mathcal{R}_1(\tau)\}/\{\tau - \mathcal{R}_0(\tau)\}$. The asymptotic variances of the ratios can be derived using the δ-method.

1.3 Combination tests

We have discussed two classes of test statistics: weighted log-rank tests and weighted Kaplan-Meier tests. A new set of test statistics may be derived by combining some members within a class and/or across the classes. This is potentially useful in the presence of non-proportional hazards. For example, the Harrington-Fleming $G^{0,1}$ will put more weight for the later time point, leading to a more powerful test when there is a late separation in the hazards. In the meanwhile, the log-rank test $G^{0,0}$ is also useful in case that the proportional hazards assumption holds. It is therefore desired to combine the two tests together. A way to do so is the so-called maximum combination tests which takes the form

$$Z_{max} = \max_{1 \le j \le J} \{Z_{W_j}\}$$

where W_j is a weight function and the Z_{W_j} is the corresponding standardized weighted log-rank statistic, $j = 1, \ldots, J$. Note that we have reverse the sign of the test statistics so that a larger value means a bigger treatment effect. A one-sided size α test can be constructed as $Pr(Z_{max} > c_\alpha) = \alpha$.

For example, Lee (2007) proposed a combination that can be sensitive to PH (FH$(0,0)$), late-separation (FH$(0,2)$), early-separation (FH$(2,0)$) and middle-separation (FH$(2,2)$).

We consider the combination of FH$(0,0)$, FH$(0,1)$, FH$(1,1)$, FH$(1,0)$ as suggested by Karrison (2016).

To conduct inference on Z_{max}, i.e. compute the critical value c_α, there are two methods that can be used. The first method is based on the computation of the correlation coefficients of the components of the combination. In fact, for any two weighted log-rank statistics $\mathcal{L}_{W_1}(\tau)$ and $\mathcal{L}_{W_2}(\tau)$, under $H_0(\tau)$ and suitable regularity conditions, the covariance of $\sqrt{n}\mathcal{L}_{W_1}(\tau)$ and $\sqrt{n}\mathcal{L}_{W_2}(\tau)$ is $C_{W_1,W_2}(\tau)$, which can be consistently estimated by

$$\mathcal{C}_{W_1,W_2}(\tau) = \int_0^\tau W_1(s)W_2(s)\frac{\Gamma_1(s)\Gamma_0(s)}{\Gamma^2(s)}dN(s).$$

As a result, the correlation coefficient between the two normalized random variables Z_{W_1} and Z_{W_2} is consistently estimated as

$$\frac{\mathcal{C}_{W_1,W_2}(\tau)}{\sqrt{\mathcal{V}_{W_1}(\tau)\mathcal{V}_{W_2}(\tau)}}.$$

Then the distribution of Z_{max} can be derived as the maximum of multivariate normal distribution with means zero, variances 1 and the correlation matrix defined above.

The other method is to use the "synthetic martingale" resampling technique proposed by Korosak and Lin (1999). To do so, suppose we want to conduct n_r times of resampling. For each $r = 1, \ldots, n_r$, generate a sequence of iid $N(0,1)$ random variables $(\xi_1, \ldots, \xi_n)$ such that they are independent of the observed data. Then, for each of the J weighted log-rank test statistics, we calculate

$$\mathcal{L}_{W_j}^{(r)}(\tau) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} W(s) \Big\{ Z_i - \frac{\Gamma_1(s)}{\Gamma(s)} \Big\} \xi_i dN_i(s),$$

$$Z_{W_j}^{(r)} = \frac{-\sqrt{n}\mathcal{L}_{W_j}^{(r)}(\tau)}{\sqrt{\mathcal{V}_{W_j}(\tau)}}, \qquad Z_{max}^{(r)} = \max_{1 \le j \le J} \{Z_{W_j}^{(r)}\}.$$

Then the distribution of Z_{max} can be approximated by the sample distribution of $\{Z_{max}^{(r)} : r = 1, \ldots, n_r\}$.

When the combination of weighted log-rank tests is used, the treatment effect estimate is taken as the estimated hazard ratio obtained from the weighted Cox model corresponding to the weighted log-rank test with the smallest p-value.

Another type of combination tests can combine the tests within the weighted Kaplan-Meier class, for example, one may combine the difference of restricted mean survival times $\int_0^{\tau} \{\widehat{S}_1(s) - \widehat{S}_0(s)\} ds$ with the landmark analysis $\widehat{S}_1(\tau-) - \widehat{S}_0(\tau-)$.

Combination tests can be constructed by combining tests from different classes. For example, one may want to combine the log-rank statistic with the difference of the restricted mean survival times.

The distribution of these combination tests can be derived using large sample theory. In practice, to account for small to moderate sample sizes, resampling methods such as bootstrapping are often used to approximate the distribution of the combination test statistics.

References

[1] Buyske S, Fagerstrom R, Ying Z. A Class of Weighted Log-Rank Tests for Survival Data When the Event is Rare. *Journal of the American Statistical Association* 2000; **95**: 249–258.

[2] Chauvel C, O'Quigley J. Tests for comparing estimated survival functions. *Biometrika* 2014; **101**: 535–552.

[3] Karrison TG. Versatile tests for comparing survival curves based on weighted log-rank statistics. *The Stata Journal* 2016; **16**: 678–690.

[4] Kosorok MR, Lin C-Y. The Versatility of Function-Indexed Weighted Log-Rank Statistics. *Journal of the American Statistical Association* 1999; **94**: 320–332.

[5] Lee S-H. On the versatility of the combination of the weighted log-rank statistics. *Computational Statistics & Data Analysis* 2007; **51**: 6557–6564.

[6] Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *JASA* 1989; **84**: 1074–1078.

[7] Lin RS, Leon LF. Estimation of treatment effects in weighted log-rank tests. *Contemporary Clinical Trials Communications* 2017; **8**: 147–155.

[8] Pepe M. Fleming T. Weighted Kaplan-Meier statistics: a class of distance tests for censored survival data. *Biometrics* 1989; **45**: 497–507.

[9] Pepe M. Fleming T. Weighted Kaplan-Meier Statistics: Large Sample and Optimality Considerations. *JRSSB* 1991; **53**: 341–352.

[10] [1993a] Sasieni P. Some new estimators for Cox regression. *Annals of Statistics* 1993; **21**: 1721–1759.

[11] [1993b] Sasieni P. Maximum Weighted Partial Likelihood Estimators for the Cox Model. *Journal of the American Statistical Association* 1993; **88**: 144–152.

[12] Uno H, Tian L, Claggett B, Wei LJ. A versatile test for equality of two survival functions based on weighted differences of Kaplan-Meier curves. *Statistics in Medicine* 2015; **34**: 3680–3695.

[13] Yang S, Prentice R. Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data. *Biometrika* 2005; **92**: 1–17.

[14] Yang S, Prentice R. Improved Logrank-Type Tests for Survival Data Using Adaptive Weights. *Biometrics* 2010; **66**: 30–38.

[15] Zhao L, Claggett B, Tian L, et al. On the restricted mean survival time curve in survival analysis. *Biometrics* 2016; **72**: 215–221.