*Harvard University*
Harvard University Biostatistics Working Paper Series

*Year* 2013 *Paper* 156

# On the Restricted Mean Event Time in Survival Analysis

Lu Tian[*] Lihui Zhao[†]
L. J. Wei[‡]

[*] Stanford University School of Medicine, lutian@stanford.edu

[†] Northwestern University, lihui.zhao@northwestern.edu

[‡] Harvard University, wei@hsph.harvard.edu

# On the restricted mean event time in survival analysis

Lu Tian, Lihui Zhao and LJ Wei

February 26, 2013

### Abstract

For designing, monitoring and analyzing a longitudinal study with an event time as the outcome variable, the restricted mean event time (RMET) is an easily interpretable, clinically meaningful summary of the survival function in the presence of censoring. The RMET is the average of all potential event times measured up to a time point $\tau$, which can be estimated consistently by the area under the Kaplan-Meier curve over $[0, \tau]$. In this paper, we present inference procedures for model-free parameters based on RMETs under the one- and two-sample settings. We then propose a new regression model, which relates the RMET to its covariates directly for predicting the subject-specific RMETs. Since the standard Cox and the accelerated failure time models can also be used for estimating such RMETs, we utilize a cross validation procedure to select the "best" working model. Lastly we draw inferences for the subject-specific RMETs based on the final candidate model using an independent data set or a "hold-out" sample from the original data set. All the proposals are illustrated with the data from the a HIV clinical trial conducted by the AIDS Clinical Trials Group and the PBC study conducted by the Mayo Clinic.

**Keywords**: Accelerated failure time model; Cox model; cross validation, hold-out sample, personalized medicine; perturbation-resampling method

# 1 Introduction

For a longitudinal study with time $T$ to a specific event as the primary outcome variable, commonly used summary measures for the distribution of $T$ are the mean, median or $t$-year event rate. Due to potential censoring for $T$, the mean may not be well estimated. If the censoring is heavy, the median cannot be identified empirically either. The $t-$year survival rate may not be suitable for summarizing the global profile of $T$. On the other hand, based on the design of the study and clinical considerations, one may pre-specify a time point $\tau$ and utilize the expected value $\mu$ of $Y = \min(T, \tau)$, the so-called restricted mean event time (RMET), as a summary parameter. This parameter is the mean of $T$ for all potential study patients followed up to time $\tau$, which has a heuristic and clinically meaningful interpretation (Irwin, 1949; Karrison, 1987; Zucker, 1998; Murray & Tsiatis, 1999; Chen & Tsiatis, 2001; Andersen et al., 2004; Zhang & Schaubel, 2011; Royston & Parmar, 2011; Zhao et al., 2012). Moreover, this model-free parameter can be estimated consistently via the standard Kaplan-Meier (KM) curve, that is, the area under the curve up to $\tau$. Inferences about the RMET can be obtained accordingly.

For a study for comparing two groups, say, $A$ and $B$, with event time observations, practitioners routinely use the hazard ratio to quantify the between-group difference. When the proportional hazards assumption is not valid, the standard maximum partial likelihood estimator of the hazard ratio approximates a parameter which is difficult, if not impossible, to interpret as the treatment contrast (Kalbfleisch & Prentice, 1981; Lin & Wei, 1989; Xu & O'Quigley, 2000; Rudser et al., 2012). Moreover, this parameter depends, oddly, on the nuisance, study-specific censoring distributions. It follows that the hazard ratio estimators at the interim and final analyses from the same study or estimators from independent studies with an identical study population would estimate different, uninterpretable parameters due to differential follow-up patterns. In fact, any model-based estimate in survival analysis may have this problem. Therefore, it is highly desirable to consider an estimable, model-free and censoring-independent parameter to quantify the treatment difference for coherent and consistent assessments between interim and final analyses within a study, as well as across independent studies. Model-free parameters for the treatment difference can be constructed via two RMETs, say, $\mu_A$ and $\mu_B$. As an example, to evaluate the added value of a potent protease inhibitor, indinavir, for HIV patients, a pivotal study ACTG 320 was conducted by the AIDS Clinical Trials Group (ACTG). This randomized, double-blind study (Hammer et al., 1997) compared a three-drug combination, indinavir, zidovudine and lamivudine, with the standard two-drug combination, zidovudine and lamivudine. There were 1156 patients enrolled for the study. One of the endpoints was the time to AIDS or death with the follow-up time about one year for each patient. Figure 1 presents the Kaplan-Meier curves for these two treatment groups. The hazard ratio estimate is 0.50 and the corresponding 0.95 confidence interval is (0.33, 0.76) with a p-value of 0.001. It is not clear if the proportional hazard assumption is valid for this study. With $\tau = 300$ days, the estimated RMET (the area under the KM curve up to 300 days) was 277 days for the control and was 288 days for the three-drug combination. The estimated difference with respect to the RMET is 11 days with the corresponding 0.95 confidence interval of (3.2, 17.3) and a p-value of 0.005. Although the treatment efficacy for the three drug combination is highly statistically significant, its clinical benefit is debatable, considering the relatively short follow-up time of the study. On the other hand, if we mimic the concept of the hazard ratio or relative risk as a summary measure for the treatment contrast, one may consider a model-free ratio $R$ of $(\tau - \mu_B)$ and $(\tau - \mu_A)$. With the above HIV data, if $B$ is the new treatment with three drug combination, the estimated $R$ is 0.55 with a p-value of $9.3 \times 10^{-6}$, also an impressive statistically significant result. For a single arm, $(\tau - \mu)$ is the average of the days lost from the healthy state up to $\tau$, a meaningful alternative to $\mu$ as a summary parameter for the distribution of $T$. Note that the above confidence interval estimates and p-values were obtained using a perturbation-resampling method detailed in Section 2.

In this paper, we first present the inference procedures for one- and two-sample problems and then consider regression models for the RMET. For the regression analysis, our goal is to build a prediction model via an extensive model selection process to stratify future patients using the patients' "baseline" covariates. The existing regression models such as the Cox model can be candidates to create such a stratification system. However, it seems more natural to model the RMET with the covariates directly, not via the hazard function (Andersen et al., 2004). In this article, we consider a new class of models which takes this approach and study the properties of the corresponding inference procedures. Since it is unlikely that any model will be precisely correct, our ultimate goal is to choose the best "fitted" model to stratify the future patients. To avoid overly optimistic results, we randomly split the data set into two pieces. Based on the first piece, called the training set, we utilize a cross validation procedure to build and select the final model. We then use the second data set, called the holdout set, to make inferences about the RMETs over a range of scores created from the final model. We use a data set from a well-known clinical study conducted at Mayo Clinic (Therneau & Grambsch, 2000) for treating a liver disease to illustrate the proposals for individualized prediction.

## 2 One- and two-sample inference procedures for RMET

For a typical subject with event time $T$, let $Z$ be the corresponding $q$-dimensional baseline covariate vector. Suppose that $T$ is subject to right censoring by a random variable $C$, which is assumed to be independent of $T$ and $Z$. The observable quantities are $(U, \Delta, Z)$, where $U = \min(T, C)$, $\Delta = I(T \leq C)$, and $I(\cdot)$ is the indicator function. The data, $\{(U_i, \Delta_i, Z_i); i = 1, \ldots, n\}$, consist of $n$ independent copies of $(U, \Delta, Z)$. Suppose that for a time point $\tau$, $\text{pr}(U \geq \tau) > 0$. The restricted survival time $Y = \min(T, \tau)$ may also be censored, but its expected value $\mu$ is estimable. Let $Y_i$ be the corresponding $Y$ for the $i$th subject, $i = 1, \ldots, n$. A natural estimator for $\mu$ is

$$\hat{\mu} = \int_0^{\tau} \hat{S}(u) du,$$

where $\hat{S}(u)$ is the KM estimator for the survival function of $T$ based on $\{(U_i, \Delta_i), i = 1, \ldots, n\}$. Alternatively, one may employ the inverse probability censoring weighting method to estimate $\mu$ as

$$\tilde{\mu} = n^{-1} \sum_{i=1}^{n} \frac{\tilde{\Delta}_i}{\hat{G}(Y_i)} Y_i,$$

where $\tilde{\Delta}_i = I(Y_i \leq C_i)$ and $\hat{G}(\cdot)$ is the KM estimator of the censoring time $C$ based on $\{(U_i, 1 - \Delta_i), i = 1, \ldots, n\}$. In Appendix A we show that $\tilde{\mu}$ and $\hat{\mu}$ are asymptotically equivalent at the root $n$ rate. Similar observations had been made by Satten & Datta (2001) with respect to KM estimator and its inverse weighting counterpart. It is straightforward to show that as $n \to \infty$, $n^{1/2}(\hat{\mu} - \mu)$ is approximately normal with mean zero and variance $\sigma^2$, which can be estimated analytically or by a perturbation-resampling method. Specifically, let

$$\mu^* = n^{-1} \sum_{i=1}^{n} \frac{\tilde{\Delta}_i}{G^*(Y_i)} Y_i Q_i,$$

where $\{Q_1, \cdots, Q_n\}$ are positive random variables with unit mean and variance and independent of the observed data and $G^*(\cdot)$ is a perturbed version of the KM estimator of the censoring variable. Here,

$$G^*(t) = \hat{G}(t) \left(1 - \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{d\hat{M}_i^C(u)}{\sum_{j=1}^{n} I(U_j \geq u)} Q_i\right),$$

$\hat{M}_i^C(u) = I(U_i \leq u)(1 - \Delta_i) - \int_0^u I(U_i \geq s) d\hat{\Lambda}_C(s)$ and $\hat{\Lambda}_C(v)$ is the Nelson-Aalen estimator for the cumulative hazard function of the censoring time $C$. Then given the data, the conditional distribution of $n^{1/2}(\mu^* - \hat{\mu})$ converges to the unconditional limiting distribution of $n^{1/2}(\hat{\mu} - \mu)$. In practice, one may generate a large number, say, $M$, of replications of $\{Q_1, \cdots, Q_n\}$ and calculate the corresponding $\mu^*$. Then the empirical distribution of $M$ generated $\mu^*$s can be used to make inference about $\mu$. This resampling technique has been used for various applications in survival analysis and in general preforms better than its analytical counterpart (Tian et al., 2005; Uno et al., 2007; Li et al., 2011). Moreover, if we are interested in a function of $\mu$, say, $g_1(\mu)$, one can utilize $M$ realizations of $g_1(\mu^*)$ for making inference. Note that for censored data, the standard bootstrapping method may generate a KM curve which is not defined at time $\tau$ and $\mu$ would not be estimable via this bootstrap sample.

For the two-sample problem, let $Z$ be $A$ or $B$ as in the Introduction. Let the corresponding means $\mu$ be denoted by $\mu_A$ and $\mu_B$ and their estimators by $\hat{\mu}_A$ and $\hat{\mu}_B$, respectively. Suppose we are interested in estimating a model-free parameter $\zeta = g_2(\mu_A, \mu_B)$, where $g_2(\cdot, \cdot)$ is a smooth bivariate function. For example, $g_2(a, b) = b - a$, $b/a$, $(\tau - a)/(\tau - b)$ or $\{b/(\tau - b)\}/\{a/(\tau - a)\}$. Then $\zeta$ can be estimated consistently with $\hat{\zeta} = g_2(\hat{\mu}_A, \hat{\mu}_B)$ and its variance estimate can be obtained via the delta-method. Alternatively, the distribution of $(\hat{\zeta} - \zeta)$ can be approximated by the conditional distribution of $(\zeta^* - \hat{\zeta})$, where $\zeta^* = g_2(\mu^*_A, \mu^*_B)$ and $\mu^*_A$ and $\mu^*_B$ are the perturbed $\mu^*$ for groups $A$ and $B$, respectively. The 0.95 confidence interval estimates and p-values for the HIV example in the Introduction were obtained with $M = 1000$ and $Q$s generated from the unit exponential.

# 3 Regression Models for RMET

If the parameter of interest is the RMET, it is natural to model

$$\mu(z) = E(Y|Z = z)$$

with the covariate vector $Z = z$ directly (Andersen et al., 2004). For example, one may consider

$$\eta\{\mu(z)\} = \beta' X, \tag{1}$$

where $\eta(\cdot)$ is a given smooth, strictly increasing function from $[0, \tau]$ to the real line, $\beta$ is a $(q + 1)$-dimension unknown vector and $X' = (1, Z')$. A special link function is $\eta(a) = \log\{a/(\tau - a)\}$, which mimics the logistic regression. Note that with this specific link, for the two sample problem, the regression coefficient of the treatment indicator is

$$\log\left\{\frac{\mu_B(\tau-\mu_A)}{\mu_A(\tau-\mu_B)}\right\},$$

an odds-ratio like summary for the group contrast, which was used for analyzing HIV data discussed in the Introduction.

For the general link function $\eta(\cdot)$, following the least squares principle, an inverse probability censoring weighted estimating function of $\beta$ is

$$S_n(\beta)=n^{-1}\sum_{i=1}^{n}\frac{\tilde{\Delta}_i}{\hat{G}(Y_i)}X_i\left\{Y_i-\eta^{-1}(\beta'X_i)\right\}.$$

Let $\hat{\beta}$ be the unique root of $S_n(\beta)=0$. Under mild regularity conditions, one can show that $S_n(\beta)$ uniformly converges to a monotone limiting function

$$S(\beta)=E\left[X\left\{\mu(Z)-\eta^{-1}(\beta'X)\right\}\right]$$

in probability. Let $\bar{\beta}$ be the root of $S(\beta)=0$. It follows that $\hat{\beta}$ converges to $\bar{\beta}$ in probability, even when the model is misspecified. As $n\rightarrow\infty$, $n^{1/2}(\hat{\beta}-\bar{\beta})$ converges weakly to a mean zero Gaussian distribution. Moreover, we can make statistical inference for $\bar{\beta}$ via the perturbation-resampling methods similar to those given in the previous section. Specifically, let $\beta^*$ be the root of the perturbed estimating equation

$$n^{-1}\sum_{i=1}^{n}\frac{\tilde{\Delta}_i}{\hat{G}^*(Y_i)}X_i\left\{Y_i-\eta^{-1}(\beta'X_i)\right\}Q_i=0.$$

It follows from the argument similar to that in Lin et al. (1993), Park & Wei (2003), and Tian et al. (2005) that the conditional distribution of $n^{1/2}(\beta^*-\hat{\beta})$ given the observed data can approximate the unconditional limiting distribution of $n^{1/2}(\hat{\beta}-\bar{\beta})$. In Appendix B, we justify this large sample approximation.

Note that Andersen et al. (2005) studied models such as (1) via a log-link $\eta(\cdot)$ using a psedo-observation technique to make inferences about the regression coefficient assuming the model is correctly specified. It can be shown that in general our estimator $\hat{\beta}$ would converge to the same parameter as that with the procedure taken by Andersen et al. (2005) when the model (1) is *correctly specified.* It would be theoretically interesting to compare the efficiency of these two procedures with respect to the estimation of the regression coefficients. However, since it is unlikely that the working model would be correctly specified, in practice it seems more relevant to evaluate such a model from the prediction point of view as we do in this article.

Using the above model, one may estimate $\mu(z)$ by $\hat{\mu}(z) = \eta^{-1}(\hat{\beta}'x)$, for any fixed $Z = z$, where $x' = (1, z')$. The distribution of $\{\hat{\mu}(z) - \eta^{-1}(\bar{\beta}'x)\}$ can be approximated by the above resampling method. Note that $\mu(z)$ can also be estimated via, for example, a Cox model (Cox, 1972). Specifically, let the hazard function for given $z$ be

$$\lambda(t|Z = z) = \lambda_0(t)e^{\gamma' z},$$

where $\gamma$ is a $q$-dimensional unknown vector and $\lambda_0(\cdot)$ is the nuisance baseline hazard function. It follows that $\mu(z)$ can be estimated by

$$\hat{\mu}(z) = \int_0^\tau \exp\{-\hat{\Lambda}_0(s)e^{\hat{\gamma}' z}\}ds,$$

where $\hat{\gamma}$ and $\hat{\Lambda}_0(s)$ are the maximum partial likelihood estimator for $\gamma$ and the Breslow estimator for $\Lambda_0(s) = \int_0^s \lambda_0(v)dv$, respectively.

Alternatively, one may use the accelerated failure time (AFT) model (Kalbfleisch & Prentice, 2002)

$$\log(T) = \gamma' Z + \epsilon,$$

to make inference about $\mu(z)$, where $\gamma$ is a $q$-dimensional unknown vector and $\epsilon$ is the error term whose distribution is entirely unspecified. Here, $\gamma$ can be estimated via a rank-based estimating function (Ritov, 1990; Tsiatis, 1990; Wei et al., 1990; Jin et al., 2003). Let $\hat{\gamma}$ be the corresponding estimator for $\gamma$. One may estimate the survival function of $e^{\epsilon}$ by KM estimator based on the data $\{(U_i e^{-\hat{\gamma}' Z_i}, \Delta_i), i = 1, \cdots, n\}$. Let the resulting estimator be denoted by $\hat{S}_0(\cdot)$. Then one can estimate $\mu(z)$ by $\hat{\mu}(z) = \int_0^{\tau} \hat{S}_0(e^{-\hat{\gamma}' z} s) ds$. Note that when $\text{pr}(C > e^{\gamma' Z} \sup e^{\epsilon}) > 0$, $\hat{\mu}(z)$ is estimable for any given covariate $z$. In practice, we can always set the censoring indicator at one for the observation with the largest $U_i e^{-\hat{\gamma}' Z_i}$ in estimating the survival function of $e^{\epsilon}$. Although these estimators for $\mu(z)$ may not be consistent and in general depend on the censoring distribution under misspecified model, they still can be reasonable predictions for the RMET.

## 4 Model Selection and Evaluation

All the models for estimating $\mu(z)$ discussed in the previous section are approximations to the true model. To compare these models, one may compare the observed restricted event time $Y$ with the covariate vector $z$ and its predicted $\hat{\mu}(z)$. A reasonable predicted error measure is $E|Y - \hat{\mu}(Z)|$, where the expected value is with respect to the data and the future subject's $(Y, Z)$. If there is no censoring, the empirical apparent prediction error is

$$n^{-1} \sum_{i=1}^{n} |Y_i - \hat{\mu}(Z_i)|,$$

which is obtained by first using the entire data to compute $\hat{\mu}(\cdot)$ and then using the same data to estimate the predicted error. Such an estimator may be biased downward (Stone, 1974; Geisser, 1975). An alternative is to utilize a cross-validation procedure to estimate such a predicted error (Tian et al., 2007; Uno et al., 2007).

Specifically, consider a class of models for $\mu(Z)$. For each model, we randomly split the data set into $K$ disjoint subsets of approximately equal sizes, denoted by $\{\mathcal{I}_k, k = 1, \ldots, K\}$. For each $k$, we use all observations which are not in $\mathcal{I}_k$ to obtain a model-based prediction rule $\hat{\mu}_{(-k)}(Z)$ for $Y$, and then estimate the total absolute prediction error for observations in $\mathcal{I}_k$ by

$$\hat{D}_k = \sum_{j \in \mathcal{I}_k} \frac{\tilde{\Delta}_j}{\hat{G}(Y_j)} \left| Y_j - \hat{\mu}_{(-k)}(Z_j) \right| .$$

Then we use the average $n^{-1} \sum_{k=1}^{K} \hat{D}_k$ as a $K$-fold cross-validation estimate for the absolute prediction error. We may repeat the aforementioned procedure a large number of, say $B$, times with different random partitions. Then the average of the resulting $B$ cross-validated estimates is the final $B$ random $K$-fold cross-validation estimate for the absolute prediction error of the fitted regression model. Generally, the model which yields the smallest cross-validated absolute prediction error estimate among all candidate models is chosen as the final model. On the other hand, a parsimonious model may be chosen if its empirical predicted error is comparable with that for the best one. We then refit the entire data set with this selected model for making predictions based on $\hat{\mu}(\cdot)$.

Note that in the training stage of this cross validation process, a candidate model may be obtained via a complex variable selection process. For example, a Cox model may be built with a stepwise regression or lasso procedure. In this case, the final choice for creating the score would be refitting the entire data set with the selected model building algorithm.

# 5 Nonparametric Inference About Subject-Specific RMET

Now, let the observed $\hat{\mu}(\cdot)$ from the final selected model be denoted by $\hat{\mu}_{opt}(\cdot)$ and for a future subject with $(Y, Z)$, let its prediction score be denoted by $V = \hat{\mu}_{opt}(Z)$. That is, for each future subject, the covariate vector $Z$ is reduced to a one-dimensional $V$ which is a function of $Z$. If the selected model is close to the true one, we expect that $E(Y|V) \approx \mu(Z) \approx V$. In general, however the group mean $\xi(v) = E(Y|V = v)$ by clustering all subjects with $Z$, whose $\hat{\mu}_{opt}(Z) = v$, may be quite different from the identity function. Therefore, the conventional parametric inferences for predicting $\xi(v)$ via the selected model may not be valid. On the other hand, since we reduce the covariate information to a univariate score $V$, one may utilize a nonparametric estimation procedure to draw valid inferences about $\xi(\cdot)$.

To make nonparametric inference about $\xi(v)$ simultaneously across a range of the score $v$, we use a fresh independent data set or "hold-out" set from the original data set. With slight abuse of notations, let such a fresh data set be denoted by $\{(U_i, \Delta_i, V_i), i = 1, \cdots, n\}$. We propose to use local linear smoothing method to estimate $\xi(v)$ nonparametrically. To this end, for a score $v$ inside the support of $V$, let $\hat{a}$ and $\hat{b}$ be the solution of the estimating equation

$$S_n(a, b; v) = \sum_{i=1}^{n} \frac{K_h(V_i - v)\tilde{\Delta}_i}{\hat{G}(Y_i|v)} \begin{pmatrix} 1 \\ V_i - v \end{pmatrix} \left[Y_i - \tilde{\eta}^{-1}\{a + b(V_i - v)\}\right] = 0,$$

where $K(\cdot)$ is a smooth symmetric kernel function with a finite support, $K_h(s) = K(s/h)/h$, $h = o_p(1)$ is the smoothing bandwidth,

$$\hat{G}(t|v) = \exp\left\{-\sum_{i=1}^{n} \int_0^t \frac{dN_i^C(u)K_h(V_i - v)}{\sum_{j=1}^{n}(U_j \geq u)K_h(V_j - v)}\right\}$$

is the local nonparametric estimator for the survival function of $C$ (Dabrowska, 1987, 1989) and $N_i^C(u) = I(\min(U_i, \tau) \leq u)(1 - \tilde{\Delta}_i)$. Here $\tilde{\eta}(\cdot)$ is a strictly increasing function from $[0, \tau]$ to the entire real line given a priori. The resulting local linear estimator for $\xi(v)$ is $\hat{\xi}(v) = \tilde{\eta}^{-1}(\hat{a})$. As $n \to \infty$ and $nh^5 = o_p(1)$, $(nh)^{1/2}\{\hat{\xi}(v) - \xi(v)\}$ converges weakly to a mean zero Gaussian. The details are given in Appendix C. Since the censoring time $C$ is assumed to be independent of $V$, generally the nonparametric KM estimator based on entire sample is used in the inverse probability weighting method for $S_n(a, b; v)$. Here, we use the local estimator $\hat{G}(t|v)$ in the above estimating equation. In the Appendix C, we show that this estimation procedure results in a more accurate estimator for $\xi(v)$ than that using $\hat{G}(\cdot)$. Note that when the empirical distribution of $\{V_i, i = 1, \cdots, m\}$ is quite non-uniform, transforming the score via an appropriate function before smoothing could potentially improve the performance of the kernel estimation (Wand et al., 1991; Park et al., 1997; Cai et al., 2010).

The aforementioned perturbation-resampling procedure in section 2 can be used to estimate the variance of $\hat{\xi}(v)$. To this end, let $\hat{a}^*$ and $\hat{b}^*$ be the solution of the perturbed estimating equation

$$S_n^*(a,b;v) = \sum_{i=1}^n Q_i \frac{K_h(V_i - v)\tilde{\Delta}_i}{G^*(Y_i|v)} \begin{pmatrix} 1 \\ V_i - v \end{pmatrix} \left[Y_i - \tilde{\eta}^{-1}\{a + b(V_i - v)\}\right] = 0,$$

where

$$G^*(t|v) = \exp\left\{-\sum_{i=1}^n \int_0^t \frac{Q_i dN_i^C(u) K_h(V_i - v)}{\sum_{j=1}^n Q_j (U_j \geq u) K_h(V_j - v)}\right\}.$$

Then a perturbed estimator for $\xi(v)$ is $\hat{\xi}^*(v) = \tilde{\eta}^{-1}(\hat{a}^*)$. Conditional on the observed data, the limiting distribution of $(nh)^{1/2}\{\hat{\xi}^*(v) - \hat{\xi}(v)\}$ approximates the unconditional counterpart of $(nh)^{1/2}\{\hat{\xi}(v) - \xi(v)\}$. It follows that one can estimate the variance of $\hat{\xi}(v)$ by $\hat{\sigma}^2(v)$, the empirical variance of $M$ realized $\hat{\xi}^*(v)'$s. Based on generated $\hat{\xi}^*(v)$, one may construct $(1-2\alpha)$ confidence interval of $\xi(v)$ as

$$\left[\hat{\xi}(v) - c_\alpha \hat{\sigma}(v), \hat{\xi}(v) + c_\alpha \hat{\sigma}(v)\right],$$

where $c_\alpha$ is the upper $100\alpha$ percentage point of the standard normal.

It is important to note that as a process, the standardized $\hat{\xi}(\cdot)$ does not converge weakly to a tight process (Bickel & Rosenblatt, 1973). On the other hand, one can use the strong approximation theory to show that the distribution of the supremum of the standardized process can still be approximated by that of the supremum of its perturbed counterpart (Gilbert et al., 2002). It follows that for an interval $[v_1, v_2]$, a subset of the support of $V$, the $(1-2\alpha)$ simultaneous confidence band of $\xi(v), v \in [v_1, v_2]$ can be constructed similarly as

$$[\hat{\xi}(v) - d_\alpha \hat{\sigma}(v), \hat{\xi}(v) + d_\alpha \hat{\sigma}(v)],$$

where

$$\text{pr}\left\{ \sup_{v\in[v_1,v_2]} \frac{|\hat{\xi}^*(v)-\hat{\xi}(v)|}{\hat{\sigma}(v)} < d_\alpha \,\middle|\, (U_i,\Delta_i,V_i), i=1,\ldots,n \right\} = 1-2\alpha,$$

where $[v_1, v_2]$ is an interval within the support of $V$. Since we require that the bandwidth $h$ converges to zero at a rate faster than the "optimal" $O_p(n^{-1/5})$ rate, the bias of $\hat{\xi}(v)$ does not play any role in the construction of the confidence interval or band for $\xi(v)$.

As with any nonparametric function estimation problem, it is crucial to choose an appropriate bandwidth $h$ in order to make proper inference about $\xi(v)$. Here, we propose a $L$ fold cross-validation procedure to choose an optimal $h$ value which minimizes a weighted cross-validated absolute prediction error. To this end, we randomly split the data set into $L$ disjoint subsets of approximately equal sizes, denoted by $\{\mathcal{I}_l, l=1,\ldots,L\}$. For any fixed $h$, we use all observations which are not in $\mathcal{I}_l$ to obtain an estimator $\hat{\xi}_{(-l)}(v,h)$ for predicting $Y$, and then estimate the total absolute prediction errors based on observations in $\mathcal{I}_l$ by

$$\hat{D}_l(h) = \sum_{j\in\mathcal{I}_l} \frac{\tilde{\Delta}_j}{\hat{G}(Y_j)} \left| Y_j - \hat{\xi}_{(-l)}(V_j,h) \right|.$$

Then we may use $\hat{D}(h) = n^{-1}\sum_{l=1}^{L}\hat{D}_l(h)$ as a final estimate for the absolute prediction error. We may choose $h_{opt}$ as the minimizer of $\hat{D}(h)$. In practice, to reduce the asymptotic bias, we propose to use a bandwidth slightly smaller than $h_{opt}$, which is expected to be in the order of $O_p(n^{-1/5})$. For example, $h = h_{opt} \times n^{-0.05}$ can be used in local smoothing estimation (Tian et al., 2005; Cai et al., 2010).

# 6 Example for Subject-Specific Prediction

In this section, we use a well-known data set from a liver study to illustrate how to build and select a model, and make inferences simultaneously about the RMETs over a range of scores created by the final model. This liver disease study in primary biliary cirrhosis (PBC) was conducted between 1974 and 1984 to evaluate the drug D-penicillamine, which was found to be futile with respect to the patient's mortality. The investigators for the study then used this rich data set to build a prediction model with respect to mortality (Fleming & Harrington, 1991). There were a total of 418 patients involved in the study including 112 patients who did not participate in the clinical trials, but had baseline and mortality information. For illustration, any missing baseline value was imputed by the corresponding sample mean calculated from its observed counterparts in the study. We randomly split the data set with equal sizes as the training and hold-out sets.

For our analysis, we consider sixteen baseline covariates: gender, histological stage of the disease (1, 2, 3, and 4), presence of ascites, edema, hepatomegaly or enlarged liver, blood vessel malformations in the skin, log-transformed age, serum albumin, alkaline phosphotase, aspartate aminotransferase, serum bilirubin, serum cholesterol, urine copper, platelet count, standardized blood clotting time and triglycerides. Three models discussed in Section 4 with these covariates included additively were considered in the model selection. They are the Cox model, the AFT model, and the new RMET model. Moreover, since a Cox model with five specific covariates (edema, age, bilirubin, albumin and standardized blood clotting time) has been established as a prediction model in the literature (Fleming & Harrington, 1991), we also considered the aforementioned three types of models with these five covariates additively in our analysis. There are, therefore, six different models were considered. Note that there was no variable selection procedure involved in the model building stage for this illustration.

Figure 2 shows the KM curve for the patients' survival with the entire data set. The patients' follow-up times are up to 13 years. Since the tail part of the KM estimate is not stable. We let $\tau = 10$ years for illustration. The overall 10-year survival rate is about 44%. Table 1 presents the $L_1$ prediction error estimates for the RMET up to 10 years for the three model building procedures based on 100 random 5-fold cross-validations. With cross-validation, the $L_1$ prediction error is minimized at 1.94 when the proposed regression model (1) with the logistic link function $\eta(\mu) = \log\{\mu/(\tau - \mu)\}$ based on five baseline covariates is utilized.

The final model is obtained by fitting the RMET model with five covariates:

$$\eta\{\mu(z)\} = 6.36 - 0.14 \times \text{edema} - 2.95 \times \log(\text{age}) + 0.99 \times \log(\text{bilirubin})$$

$$-0.35 \times \log(\text{albumin}) + 0.34 \times \log(\text{clotting time}).$$

We then use the score created by this model to make prediction and stratification for subjects in the hold-out set.

For predicting future restricted event time, we use the procedures proposed in Section 5 to estimate the subject-specific RMET $\xi(v)$ over a range of score $v$'s, and construct its 0.95 pointwise and simultaneous confidence intervals over the interval $[0.07, 9.52]$, where 0.07 and 9.52 are the 2nd and 98th percentiles of observed scores in the holdout set. Here, we let $K(\cdot)$ be the Epanechnikov kernel and the bandwidth be 2.1, as selected via cross-validation. The results are presented in the first panel of Figure 3. For comparison, we also present the corresponding results in the second panel of Figure 3 with the survival function of the censoring time $C$ being estimated based on the entire sample rather than locally as proposed in Section 5. As expected, the resulting estimator for $\xi(v)$ is less accurate, e.g., the 95% confidence interval for $\xi(5)$ is 24.8% wider when the survival function of $C$ is estimated based on the entire sample.

As a conventional practice, we may stratify the subjects in the hold-out set into groups such as low, intermediate and high risk groups by discretizing the continuous score. For example, we may create four classes based on the quartiles of the scores. Figure 4 presents the KM curves for these four strata. Visually these curves appear quite different. Moreover, their estimated RMETs and the standard error estimates (in paratheses) are 3.59 (0.46), 6.26 (0.53), 8.50 (0.40), and 9.14 (0.31) in years, respectively. These indicate that the scoring system does have reasonable discriminating capability with respect to the patients' RMET. How to construct an "efficient" categorization of the existing scoring system warrants future research.

# 7 Remarks

In comparing two groups with censored event time data, the point and interval estimates of the two RMETs and their counterparts for the group contrast provide much more clinically relevant information than, for example, the hazard ratio estimate. The results from the HIV data set from ACTG 320 discussed in the Introduction is a good example, in that the three drug combination is statistically significantly better than the conventional therapy, but the gain from the new treatment with respect to RMET was not as impressive from a clinical standpoint, likely due to the relatively short follow-up time. Note that for this case, the median event time cannot be estimated empirically due to heavy censoring. Moreover, we cannot evaluate models using the individual predicted error, such as the $L_1$ distance function, with the median event time. It follows that the RMET is probably the most meaningful, model-free, global measure for the distribution of the event time to evaluate the treatment efficacy. The choice of $\tau$ to define the RMET is crucial, which may be determined at the study design stage with respect to clinical relevance and feasibility of conducting the study.

Note that one of the attractive features of the model which directly relates the RMET to its covariates proposed here is that the score created is free of the censoring distribution even when the model is not correctly specified. On the other hand, those scores built from the Cox or AFT models depend on the study-specific censoring distribution when the model is misspecified.

# Appendix A: The asymptotic equivalence of $\hat{\mu}$ and $\tilde{\mu}$

Firstly, for $t \in [0, \tau]$, the KM estimator $\hat{S}(t)$ is the solution to the forward integral equation

$$S(t) = 1 - \int_0^t S(u^-) d\hat{\Lambda}(u)$$

and thus

$$
\begin{aligned}
\hat{S}(t) &= 1 - \int_0^t \hat{S}(u^-)\frac{dN(u)}{Y(u)} = 1 - \frac{1}{n}\int_0^t \frac{dN(u)}{\hat{G}_{(}u^-)} + o_p(n^{-1/2}) \\
&= 1 - \frac{1}{n}\sum_{i=1}^n \frac{I(Y_i \le t)\tilde{\Delta}_i}{\hat{G}(Y_i)} + o_p(n^{-1/2}) = \frac{1}{n}\sum_{i=1}^n \frac{I(Y_i > t)\tilde{\Delta}_i}{\hat{G}(Y_i)} + o_p(n^{-1/2})
\end{aligned}
$$

where $N(t) = \sum_{i=1}^n N_i(t)$, $Y(t) = \sum_{i=1}^n I(U_i \ge t)$ and $\hat{\Lambda}(\cdot)$ is the Nelsen-Aalen estimator of the cumulative hazard function of $Y$. In the above derivation, we have used the fact that

$$
n^{-1}Y(t) = \hat{G}(t^-)\hat{S}(t^-) + o_p(n^{-1}).
$$

Therefore

$$
\begin{aligned}
\hat{\mu} - \tilde{\mu} &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau \frac{I(Y_i > t)\tilde{\Delta}_i}{\hat{G}(Y_i)} dt - \tilde{\mu} + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^n \frac{\tilde{\Delta}_i}{\hat{G}(Y_i)} Y_i - \tilde{\mu} + o_p(n^{-1/2}) = o_p(n^{-1/2}).
\end{aligned}
$$

## Appendix B: The asymptotic properties of $\hat{\beta}$

It can be shown that $S(\beta)$ has an unique solution at $\bar{\beta}$ when model (1) is correctly specified. Let the root of the estimating equation $S_n(\beta) = 0$ be denoted by $\hat{\beta}$, $\hat{\beta}$ converges to $\bar{\beta}$ in probability. Furthermore, one can show that

$$
n^{1/2}S_n(\bar{\beta}) = n^{-1/2}\sum_{i=1}^n \kappa_i + o_p(1)
$$

where

$$
\kappa_i = X_i\left\{Y_i - \eta^{-1}(\bar{\beta}'X_i)\right\} - \int_0^\tau \frac{dM_i^C(u)}{G(u)}\left[X_i\left\{Y_i - \eta^{-1}(\bar{\beta}'X_i)\right\} - K(\bar{\beta}, u)\right],
$$

and

$$K(\bar{\beta}, u) = S(u)^{-1} E\left[X\left\{Y - \eta^{-1}(\bar{\beta}' X)\right\} I(Y \geq u)\right].$$

Coupled with the local linearity of the estimating function, this expansion implies that $n^{1/2}(\hat{\beta} - \bar{\beta})$ converges weakly to a mean zero Gaussian distribution with a variance-covariance matrix of $A^{-1}BA^{-1}$, where

$$A = E\left\{X^{\otimes 2} \dot{\eta}^{-1}(\bar{\beta}' X)\right\},$$

$$B = E\left[X^{\otimes 2}\left\{Y - \eta^{-1}(\bar{\beta}' X)\right\}^2\right]$$

$$+ \int_0^\tau E\left[X\left\{Y - \eta^{-1}(\bar{\beta}' X)\right\} - K(\bar{\beta}, u)\right]^{\otimes 2} I(T \geq u) \frac{d\Lambda_C(u)}{G(u)},$$

$\dot{\eta}^{-1}(\cdot)$ is the derivative of $\eta^{-1}(\cdot)$ and $a^{\otimes 2} = aa'$ for vector $a$.

Furthermore, one can show that

$$n^{1/2}(\beta^* - \hat{\beta}) = A^{-1} n^{1/2} S^*(\hat{\beta}) + o_{\tilde{P}}(1) = n^{-1/2} \sum_{i=1}^n (Q_i - 1)\hat{\kappa}_i + o_{\tilde{P}}(1)$$

where the probability measure $\tilde{P}$ is the product measure generated by $\{Q_1, \ldots, Q_n\}$ and data,

$$\hat{\kappa}_i = X_i\left\{Y_i - \eta^{-1}(\hat{\beta}' X_i)\right\} - n^{-1/2} \int_0^\tau \frac{d\hat{M}_i^C(u)}{\hat{G}(u)}\left[X_i\left\{Y_i - \eta^{-1}(\hat{\beta}' X_i)\right\} - \hat{K}(\hat{\beta}, u)\right] + o_p(1),$$

$\hat{M}_i^C(u) = I(U_i \leq u)(1 - \Delta_i) - \int_0^u I(U_i \geq t) d\hat{\Lambda}_C(t)$ and

$$\hat{K}(\beta, u) = \frac{\sum_{i=1}^n X_i\left\{Y_i - \eta^{-1}(\beta' X_i)\right\} I(U_i \geq u)}{\sum_{i=1}^n I(U_i \geq u)}.$$

Therefore, as $n \to \infty$ the conditional variance of $n^{1/2}(\beta^* - \hat{\beta})$ converges to

$$A^{-1}(n^{-1} \sum_{i=1}^n \hat{\kappa}_i^{\otimes 2}) A^{-1},$$

which is a consistent estimator of $A^{-1}BA^{-1}$.

## Appendix C: The Asymptotic Properties of $\hat{\xi}(v)$

Let $bh = \tilde{b}$. As $n \to \infty$, the estimating equation $S_n(a, \tilde{b}; v)$ converges to

$$S(a, \tilde{b}; v) = f_V(v) \begin{pmatrix} \left[\xi(v) - \int K(x)\tilde{\eta}^{-1}(a + \tilde{b}x)dx\right] \\ -\int xK(x)\tilde{\eta}^{-1}(a + \tilde{b}x)dx \end{pmatrix},$$

where $f_V(\cdot)$ is the density function of $V$. Since $S(a, \tilde{b}; v) = 0$ has an unique root $(a, \tilde{b}) = (\tilde{\eta}\{\xi(v)\}, 0)$, $\hat{a} \to \tilde{\eta}\{\xi(v)\}$ and $h\hat{b} \to 0$ in probability as $n \to \infty$. Furthermore, by Taylor series expansion, we have

$$\begin{aligned} &(nh)^{1/2}\{\tilde{\eta}^{-1}(\hat{a}) - \xi(v)\} \\ =&\frac{n^{1/2}}{f_V(v)h^{1/2}} \sum_{i=1}^{n} \frac{\tilde{\Delta}_i}{\hat{G}(t|v)} K_h(V_i - v) \left[Y_i - \tilde{\eta}^{-1}\{a_0 + b_0(V_i - v)\}\right] \\ =&\frac{h^{1/2}}{n^{1/2}} \sum_{i=1}^{n} \left(K_h(V_i - v)\{Y_i - \xi(V_i)\} - \int_0^\infty \frac{dM_i^C(u)}{G(u)} K_h(V_i - v)\left[\{Y_i - \xi(V_i)\} - R(u, v)\right]\right) + o_p(1) \end{aligned}$$

where

$$R(u, v) = \frac{\gamma_1(u|v)}{S_T(u|v)} \quad \text{and} \quad \gamma_j(u|v) = E[I(Y \geq u)\{Y - \xi(v)\}^j | V = v].$$

Therefore, by martingale central limit theorem $(nh)^{1/2}\{\tilde{\eta}^{-1}(\hat{a}) - \xi(v)\}$ converges weakly to a mean zero Guassian with variance

$$\nu_2 \left[ E[\{Y - \xi(v)\}^2 | V = v] + \int_0^\infty \left\{\gamma_2(u|v) - \frac{\gamma_1(u|v)^2}{S_T(u|v)}\right\} \frac{d\Lambda_C(u)}{G(u)} \right],$$

where $\nu_2 = \int K(v)^2 dv$. If we use the KM estimator $\hat{G}(u)$ to replace $\hat{G}(u|v)$ in the estimating function $S_n(a, b; v)$ and denote the corresponding estimator by $\tilde{\xi}(v)$, then one can similarly show that the variance of the $(nh)^{1/2}\{\tilde{\xi}(v) - \xi(v)\}$ is

$$\nu_2 \left[ E[\{Y - \xi(v)\}^2 | V = v] + \int_0^\infty \gamma_2(u|v) \frac{d\Lambda_C(u)}{G(u)} \right]$$

which is greater than that of $(nh)^{1/2}\{\hat{\xi}(v) - \xi(v)\}$.

# References

Andersen, P., Hansen, M. & Klein, J. (2004). Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Analysis* **10**, 335–350.

Bickel, P. & Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Annals of Statistics* **1**, 1071–1095.

Cai, T., Tian, L., Uno, H., Solomon, S. & Wei, L. (2010). Calibrating parametric subject-specific risk estimation. *Biometrika* **97**, 389–404.

Chen, P. & Tsiatis, A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics* **57**, 1030–1038.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–220.

Dabrowska, D. (1987). Non-parametric regression with censored survival time data. *Scandinavian Journal of Statistics* , 181–197.

Dabrowska, D. (1989). Uniform consistency of the kernel conditional kaplan-meier estimate. *The Annals of Statistics* **17**, 1157–1167.

Fleming, T. & Harrington, D. (1991). *Counting processes and survival analysis*, vol. 8. Wiley Online Library.

GEISSER, S. (1975). The predictive sample reuse method with applications. *Journal of American Statistical Association* **70**, 320–328.

GILBERT, P., WEI, L., KOSOROK, M. & CLEMENS, J. (2002). Simultaneous inferences on the contrast of two hazard functions with censored observations. *Biometrics* **58**, 773–780.

HAMMER, S., SQUIRES, K., HUGHES, M., GRIMES, J., DEMETER, L., CURRIER, J., ERON, J., FEINBERG, J., BALFOUR, H., DEYTON, L. et al. (1997). A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and cd4 cell counts of 200 per cubic millimeter or less. *New England Journal of Medicine-Unbound Volume* **337**, 725–733.

IRWIN, J. (1949). The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. *Journal of Hygiene* **47**, 188–189.

JIN, Z., LIN, D., WEI, L. & YING, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–353.

KALBFLEISCH, J. D. & PRENTICE, R. L. (1981). Estimation of the average hazard ratio. *Biometrika* **68**, 105–112.

KALBFLEISCH, J. D. & PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data.* New York: JohnWiley & Sons.

KARRISON, T. (1987). Restricted mean life with adjustment for covariates. *Journal of the American Statistical Association* **82**, 1169–1176.

LI, Y., TIAN, L. & WEI, L. (2011). Estimating subject-specific dependent competing risk profile with censored event time observations. *Biometrics* **67**, 427–435.

LIN, D., WEI, L. & YING, Z. (1993). Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–572.

Lin, D. Y. & Wei, L. J. (1989). The robust inference for the cox proportional hazards model. *Journal of American Statistical Association* **84**, 1074–1078.

Murray, S. & Tsiatis, A. (1999). Sequential methods for comparing years of life saved in the two-sample censored data. *Biometrics* **55**, 1085–1092.

Park, B., Kim, W., Ruppert, D., Jones, M., Signorini, D. & Kohn, R. (1997). Simple transformation techniques for improved non-parametric regression. *Scandinavian journal of statistics* **24**, 145–163.

Park, Y. & Wei, L. (2003). Estimating subject-specific survival functions under the accelerated failure time model. *Biometrika* **90**, 717–723.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *The Annals of Statistics* , 303–328.

Royston, P. & Parmar, M. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in medicine* **30**, 2409–2421.

Rudser, K., LeBlanc, M. & Emerson, S. (2012). Distribution-free inference on contrasts of arbitrary summary measures of survival. *Statistics in Medicine* **31**, 1722–1737.

Satten, G. & Datta, S. (2001). The kaplan-meier estimator as an inverse-probablity-of-censoring weighted average. *The American Statistician* **55**, 207–210.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of Royal Statitical Society. (Series B)* **36**, 111–147.

Therneau, T. & Grambsch, P. (2000). *Modeling survival data: extending the Cox model.* Springer.

Tian, L., Cai, T., Goetghebeur, E. & Wei, L. (2007). Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika* **94**, 297–311.

Tian, L., Zucker, D. & Wei, L. (2005). On the cox model with time-varying regression coefficients. *Journal of the American statistical Association* **100**, 172–183.

Tsiatis, A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics* **18**, 354–372.

Uno, H., Cai, T., Tian, L. & Wei, L. (2007). Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association* **102**, 527–537.

Wand, M., Marron, J. & Ruppert, D. (1991). Transformations in density estimation. *Journal of the American Statistical Association* **86**, 343–353.

Wei, L., Ying, Z. & Lin, D. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–851.

Xu, R. & O'Quigley, J. (2000). Estimating average regression effect under non-proportional hazards. *Biostatistics* **1**, 423–439.

Zhang, M. & Schaubel, D. (2011). Estimating differences in restricted mean lifetime using observational data subject to dependent censoring. *Biometrics* **67**, 740–749.

Zhao, L., Tian, L., Uno, H., Solomon, S., Pfeffer, M., Schindler, J. & Wei, L. (2012). Utilizing the integrated difference of two survival functions to quantify the treatment contrast for designing, monitoring, and analyzing a comparative clinical study. *Clinical Trials* **9**, 570–577.

Zucker, D. (1998). Restricted mean life with covariates: modification and extension of a useful survival analysis method. *Journal of the American Statistical Association* **93**, 702–709.

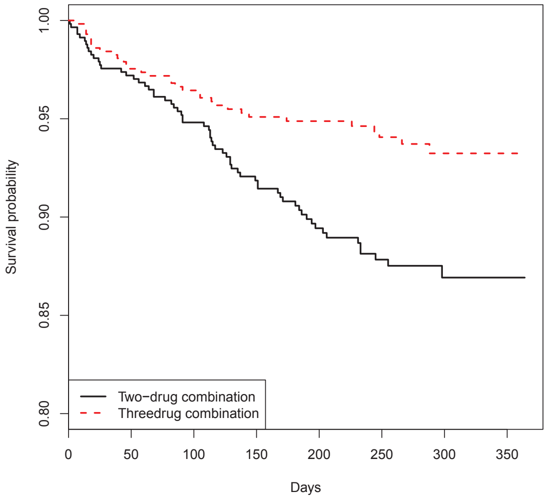

Figure 1: Kaplan-Meier estimates of the survival functions of the two randomized groups based on the ACTG 320 data

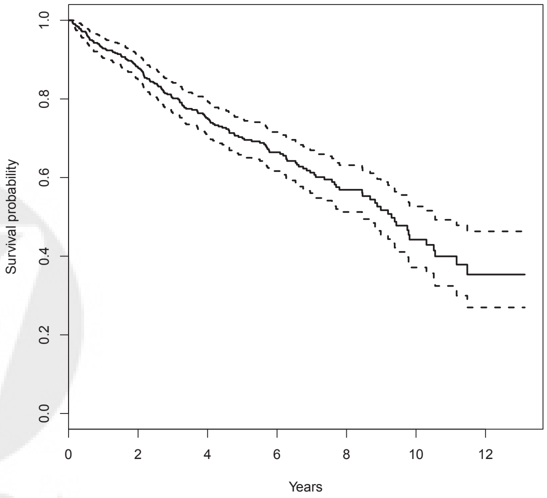

Figure 2: Kaplan-Meier estimate of the overall patient survival function based on the PBC data (Solid: point estimate, Dashed: 95% pointwise confidence interval

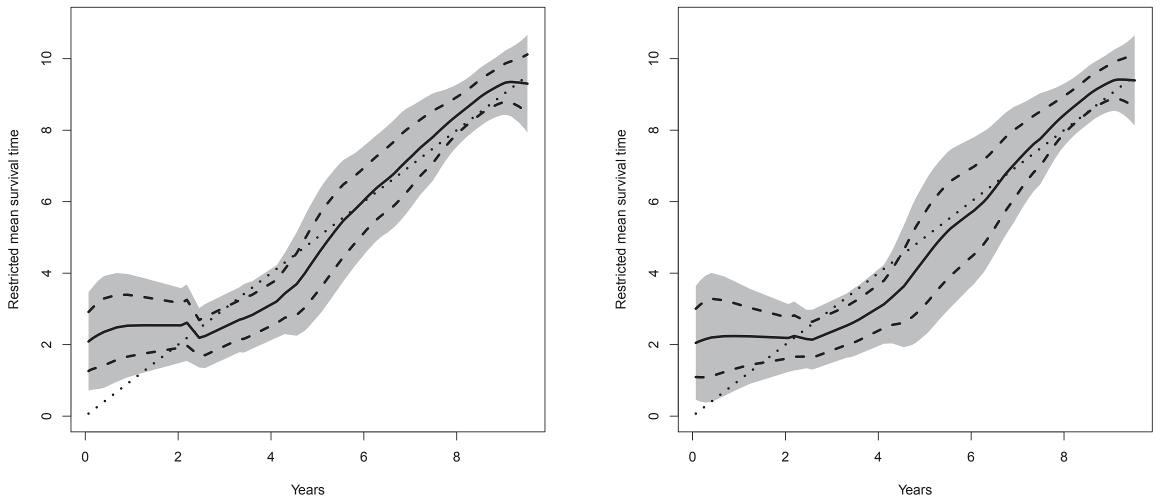

Figure 3: Estimated subject-specific restricted mean survival time (solid curve) over the score, and its 95% pointwise (dashed curve) and simultaneous confidence intervals (shaded region). The dotted line is the 45 degree reference line. The survival function of the censoring time C is estimated locally in the first panel and based on the entire sample in the second panel.

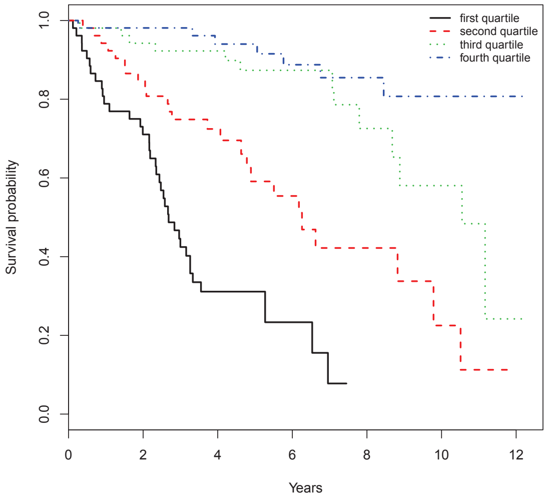

Figure 4: Kaplan-Meier estimates of the survival functions of the four strata divided by quartiles of the scores based on the PBC data

Table 1: $L_1$ prediction error estimates for the restricted mean event time up to 10 years of the three model building procedures based on 100 random 5-fold cross-validations (CV)

| | $L_1$ prediction error with CV | | |
|---|---|---|---|
| | Cox model | AFT model | New model |
| 5 covariates | 2.34 | 2.00 | 1.94 |
| 16 covariates | 2.34 | 2.11 | 2.10 |