Published in final edited form as:

Lifetime Data Anal. 2016 April ; 22(2): 299–319. doi:10.1007/s10985-015-9332-1.

On Estimation of Covariate-Specific Residual Time Quantiles under the Proportional Hazards Model

Luis Alexander Crouch,

Department of Biostatistics, University of Washington, Seattle, WA 98105, Tel.: +1-425-954-5492, crouch@uw.edu

Susanne May, and

Department of Biostatistics, University of Washington, Seattle, WA 98105

Ying Qing Chen

Program in Biostatistics and Biomathematics, Fred Hutchinson Cancer Research Center, Seattle, WA 98109

Abstract

Estimation and inference in time-to-event analysis typically focus on hazard functions and their ratios under the Cox proportional hazards model. These hazard functions, while popular in the statistical literature, are not always easily or intuitively communicated in clinical practice, such as in the settings of patient counseling or resource planning. Expressing and comparing quantiles of event times may allow for easier understanding. In this article we focus on residual time, i.e., the remaining time-to-event at an arbitrary time t given that the event has yet to occur by t . In particular, we develop estimation and inference procedures for covariate-specific quantiles of the residual time under the Cox model. Our methods and theory are assessed by simulations, and demonstrated in analysis of two real data sets.

Keywords

Censoring; Hazard function; Time-to-event

1 Introduction

Analysis of censored time-to-event data is typically centered around fitting a Cox proportional hazards model. The Cox model offers many advantages: the ability to handle censored data, no need for parametric assumptions about the baseline hazard function, interpretable output (in the form of hazard ratios), and readily available software for implementation. For example, we can consider the oropharynx carcinoma data from Kalbfleisch and Prentice (2002). In this study, 195 patients with squamous cell carcinoma were randomly assigned treatment with radiation therapy alone or together with a chemotherapeutic agent. When a Cox model is fit to assess the association between tumor (T) staging and survival time, the hazard ratios are 1.24 and 2.01 for T stages III and IV, respectively, compared to stage I/II.

Although hazard functions and their ratios are very useful, the fact that they are expressed in a conditional probabilistic likelihood can make them less intuitive to medical practitioners and patients in a clinical setting. Summary measures on the same scale as actual event times may be more easily understood. In particular, estimates of event times typical of a patient with certain characteristics (i.e., covariate-specific) provide a compelling summary measure. These estimates may be more useful in, for example, clinical counseling and resource planning. In the oropharynx carcinoma example, the median event (survival) times for T stages I/II, III, and IV are 661, 465, and 317 days, respectively, clearly highlighting the differences in survival prognosis associated with the different stages of cancer.

Estimating mean event times or quantiles of event times is not entirely novel. Dabrowska and Doksum (1987) developed estimators for mean event times and quantiles of event times under the Cox model. The application of said estimates may be clinically limited, however, by their beginning at a “time zero,” such as enrollment time, which is often subjectively chosen. A possible example of such a limitation is the median survival time for a patient diagnosed with a disease with high initial hazard that drops over time. The initial median survival time may not be relevant when considering said patient’s outlook after the initial spike in hazard. It may therefore be useful to consider the residual time: the remaining time-to-event at any given time $t > 0$, assuming the event has yet to occur by t .

Estimation of residual time has been approached in many ways. Much work has centered around calculating mean or expected residual time; as in other settings, these values are typically more statistically tractable. Other work has focused on estimating quantiles of residual time, particularly as this may allow for reduced concern about censoring. In either case, summary statistics for residual time may be based on direct modeling or on transformation of other estimated quantities, such as using estimated hazard ratios and baseline hazards from the Cox model.

Many methods for estimating residual time quantiles are generalizations of those for estimating survival time quantiles. Early estimators for survival time quantiles were based on empirical functions (Cheng, 1984), typically using the inverse of the product limit estimator proposed by Kaplan and Meier (1958). While still focusing on median survival times, Ying et al. (1995) and Yang (1999) proposed regression frameworks allowing for more general estimation of covariate effects, as opposed to simply comparing two groups’ median survival times. A different regression model, using a Bayesian approach, was formulated by Gelfand and Kottas (2003). Methods for residual time quantiles have seen less development. Recently, Jeong et al. (2008) proposed nonparametric methods for estimating median residual time based on inverting Kaplan-Meier estimators. In a follow-up paper, Jung et al. (2009) provide a regression model that allows for modeling of covariate effects on general quantile residual time. A different regression model is proposed by Ma and Yin (2010), allowing for estimation of quantiles of residual times in addition to covariate effects on them.

In this article, we aim to develop an estimator for residual time quantiles under the Cox model. As shown in Section 2, the newly developed estimator allows estimation of covariate-specific residual time quantiles. We demonstrate our estimator’s consistency,

determine its limiting distribution, and provide a consistent estimator for its variance. Also included are discussions of methods for obtaining confidence intervals and bands that do not rely on direct estimation of the variance. We further develop our method by determining the limiting distribution for a difference between two estimators of covariate-specific residual time quantiles, allowing formal comparison of residual time quantiles. In Section 3 we demonstrate our estimator's performance on simulated data. Additionally, we apply our method to two real data sets: the VA lung cancer data as presented in Kalbfleisch and Prentice (2002) and the HIVNET 012 data as presented in Jackson et al. (2003). Directions for future work are discussed in Section 4.

2 Methods for Model-based Residual Time Quantiles

Assume that T is a positive random variable representing a study subject's time-to-event, and $\mathbf{Z} \in \mathcal{R}^p$ is the p -dimensional vector of associated covariates. For $0 < q < 1$, at a given time $t > 0$, the $(1 - q) \times 100\%$ ile residual time of T is the $(1-q) \times 100\%$ ile of $T - t$ given that $T > t$. Denoted by $\theta(t, q|\mathbf{Z})$, the $(1-q) \times 100\%$ ile residual time is determined by

$$S\{t + \theta(t, q|\mathbf{Z})|\mathbf{Z}\} = q \cdot S(t|\mathbf{Z}), \quad (1)$$

where $S(t|\mathbf{Z}) = \text{pr}\{T > t | \mathbf{Z}\}$ is T 's survival function. In particular, when $q = 0.5$, it is the median residual time for $\mathbf{Z}$.

In (1), the covariates $\mathbf{Z}$ can be fairly general for the event time T . For example, if T is time to lung cancer diagnosis, $\mathbf{Z}$ can include a binary indicator of smoking-versus-not, a continuous variable of age-started-smoking, both, or some combination of other variables. In practice, to model the association between $\mathbf{Z}$ and T , a Cox proportional hazards model is usually assumed:

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}), \quad (2)$$

where $\lambda_0(\cdot) > 0$ is an unspecified baseline hazard function, $\lambda(\cdot | \mathbf{Z})$ is the hazard function associated with covariates $\mathbf{Z}$, $\boldsymbol{\beta} \in \mathcal{B} \subset \mathcal{R}^p$ is the p -dimensional regression parameter, and T denotes vector (matrix) transpose.

Under the Cox model (2), the $(1 - q) \times 100\%$ ile residual time associated with covariates $\mathbf{Z}$ at a given $t > 0$ is then

$$\theta(t, q|\mathbf{Z}) = \Lambda_0^{-1}\{\Lambda_0(t) - \log q \cdot \exp(-\boldsymbol{\beta}^T \mathbf{Z})\} - t, \quad (3)$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ is the monotonically increasing cumulative hazard function of $\lambda_0(t)$, and $\Lambda_0^{-1}(t) = \inf\{t: \Lambda_0(t) \geq s\}$ is the inverse function of $\Lambda_0(t)$. It is straightforward to show that the model-based $\theta(t, q|\mathbf{Z})$ in (3) has the following properties:

Property 1 $\theta(t, q|\mathbf{Z})$ is a monotonically decreasing function of q and $\boldsymbol{\beta}^T \mathbf{Z}$.

Property 2 When T is exponentially distributed, $\theta(t, q|\mathbf{Z})$ is constant regardless of any given $t > 0$.

These two properties capture some characteristics of $\theta(t, q|\mathbf{Z})$. For example, individuals with greater values for covariates associated with increased hazard will have lesser residual times. Although it is difficult to characterize the dependence of $\theta(t, q|\mathbf{Z})$ on t , when T is “memoryless” as with an exponential distribution, $\theta(t, q|\mathbf{Z})$ also reflects the memoryless nature by being constant. Examples for distributions of T other than exponential, such as Weibull and log-normal distributions, are given in the Online Resource Appendix and may assist in more intuitive understanding of $\theta(t, q|\mathbf{Z})$.

Now assume that a study is conducted and we would like to estimate $\theta(t, q|\mathbf{Z})$ based on the collected data. Suppose the data are collected in the form of $(X_i, \delta_i, \mathbf{Z}_i)$ for $i = 1, \dots, n$, with n being the sample size. For the i^{th} study participant, $X_i = \min(T_i, C_i)$, where T_i is failure time and C_i is censoring time; $\delta_i = I(T_i \leq C_i)$; and $\mathbf{Z}_i$ is a vector of covariates. Given $\mathbf{Z}_i$, T_i and C_i are assumed to be independent. For the collected data, we first estimate β under the Cox model (2) by the usual maximum partial likelihood method maximizing

$$\prod_{i=1}^n \left\{ \frac{\exp(\beta^T \mathbf{Z}_i)}{\sum_{j=1}^n I(X_j \geq X_i) \exp(\beta^T \mathbf{Z}_j)} \right\}^{\Delta_i}$$

with respect to β . That is, the partial likelihood estimator $\hat{\beta}$ will be solved such that

$$\sum_{i=1}^n \int_0^\infty \left\{ \mathbf{Z}_i - \frac{\mathbf{S}^{(1)}(\hat{\beta}, t)}{\mathbf{S}^{(0)}(\hat{\beta}, t)} \right\} dN_i(t) = 0,$$

where $N_i(t) = I(X_i \leq t, \delta_i = 1)$, $Y_i(t) = I(X_i \geq t)$, $\mathbf{S}^{(0)}(\beta, t) = \sum_{i=1}^n Y_i(t) \times e^{\beta^T \mathbf{Z}_i}$, and $\mathbf{S}^{(1)}(\beta, t) = \sum_{i=1}^n Y_i(t) \mathbf{Z}_i e^{\beta^T \mathbf{Z}_i}$, as in Fleming and Harrington (2005). We estimate the baseline cumulative hazard function $\Lambda_0(\cdot)$ with the Breslow estimator,

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s) \exp(\hat{\beta}^T \mathbf{Z}_i)}.$$

With $\hat{\beta}$ and $\hat{\Lambda}_0(\cdot|\mathbf{Z})$, we can in turn estimate $\theta(t, q|\mathbf{Z})$ in (3) by

$$\hat{\theta}(t, q|\mathbf{Z}) = \hat{\Lambda}_0^{-1} \{ \hat{\Lambda}_0(t) - \log q \exp(-\hat{\beta}^T \mathbf{Z}) \} - t, \quad (4)$$

where $\hat{\Lambda}_0^{-1}(s) = \inf\{t: \hat{\Lambda}_0(t) \geq s\}$. Note that $\hat{\theta}(t, q|\mathbf{Z})$ is only defined when $\hat{\Lambda}_0(t) - \log q \exp(-\hat{\beta}^T \mathbf{Z}) \leq \hat{\Lambda}_0(T^*)$ where T^* is the largest observed failure time.

For the model-based estimator $\hat{\theta}(t, q|\mathbf{Z})$ in (4), we can establish the following asymptotic properties, as summarized in

Theorem 1

We assume conditions A–D (see Appendix) of Andersen and Gill (1982) hold. If, as previously developed, $\hat{\theta}(t, q|\mathbf{Z}) = \hat{\Lambda}_0^{-1} \{ \hat{\Lambda}_0(t) - \log q \exp(-\hat{\beta}^T \mathbf{Z}) \} - t$, then as $n \rightarrow \infty$, for a given $\mathbf{Z}$,

1. $\hat{\theta}(t, q|\mathbf{Z}) \rightarrow_p \theta(t, q|\mathbf{Z})$;
2. $\sqrt{n} \{ \hat{\theta}(t, q|\mathbf{Z}) - \theta(t, q|\mathbf{Z}) \} \rightarrow_d N(0, V(t))$, where

$$V(t) = \frac{1}{\lambda_0 \{ \theta(t, q|\mathbf{Z}) + t \}^2} \left[\mathbf{A}^T \{ \boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau) \}^{-1} \mathbf{A} + \int_0^{\theta(t, q|\mathbf{Z})} \frac{\lambda_0(t+v)}{s^{(0)}(\boldsymbol{\beta}, t+v)} dv \right],$$

$$\mathbf{A} = \left[\log q \times \mathbf{Z} \exp(-\boldsymbol{\beta}^T \mathbf{Z}) + \int_0^{\theta(t, q|\mathbf{Z})} \frac{s^{(1)}(\boldsymbol{\beta}, t+v)}{s^{(0)}(\boldsymbol{\beta}, t+v)} \lambda_0(t+v) d(t+v) \right]^T,$$

and τ is such that $\int_0^\tau \lambda(t) dt < \infty$;

3. $\hat{V}(t) \rightarrow_p V(t)$, where

$$\hat{V}(t) = \frac{1}{\hat{\lambda}_0(\hat{\theta}(t, q|\mathbf{Z}) + t)^2} \left[\hat{\mathbf{A}}^T \{ \hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}}, \tau) \}^{-1} \hat{\mathbf{A}} + \int_0^{\hat{\theta}(t, q|\mathbf{Z})} \frac{d\hat{\Lambda}_0(t+v)}{S^{(0)}(\hat{\boldsymbol{\beta}}, t+v)} \right]$$

and

$$\hat{\mathbf{A}} = \left[\log q \times \mathbf{Z} \exp(-\hat{\boldsymbol{\beta}}^T \mathbf{Z}) + \int_0^{\hat{\theta}(t, q|\mathbf{Z})} \frac{S^{(1)}(\hat{\boldsymbol{\beta}}, t+v)}{\{S^{(0)}(\hat{\boldsymbol{\beta}}, t+v)\}^2} d\hat{N}(t+v) \right]^T.$$

Here, $S^{(k)}(\boldsymbol{\beta}, t) = \sum_{i=1}^n Y_i(t) \mathbf{Z}_i^{\otimes k} e^{\boldsymbol{\beta}' \mathbf{Z}_i} \rightarrow_p s^{(k)}(\boldsymbol{\beta}, t)$, for $k = 0, 1, 2$; $\mathbf{E}(\boldsymbol{\beta}, t) = S^{(1)}(\boldsymbol{\beta}, t)/S^{(0)}(\boldsymbol{\beta}, t) \rightarrow_p \mathbf{e}(\boldsymbol{\beta}, t) = s^{(1)}(\boldsymbol{\beta}, t)/s^{(0)}(\boldsymbol{\beta}, t)$, $\mathbf{V}(\boldsymbol{\beta}, t) = S^{(2)}(\boldsymbol{\beta}, t)/S^{(0)}(\boldsymbol{\beta}, t) - \{S^{(1)}(\boldsymbol{\beta}, t)/S^{(0)}(\boldsymbol{\beta}, t)\}^{\otimes 2} \rightarrow_p \mathbf{v}(\boldsymbol{\beta}, t) = s^{(2)}(\boldsymbol{\beta}, t)/s^{(0)}(\boldsymbol{\beta}, t) - \mathbf{e}(\boldsymbol{\beta}, t)^{\otimes 2}$, and

$$\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}}, \tau) = n^{-1} \int_0^\tau \sum_{i=1}^n \mathbf{V}(\hat{\boldsymbol{\beta}}, v) dN_i(v).$$

Proof of Theorem 1 is in the Appendix. Note that the estimate for variance, $\hat{V}(t)$, requires estimation of the hazard function, $\lambda_0(\cdot)$. We use a kernel smoothed estimator employing the Epanechnikov kernel (Epanechnikov, 1969) as described in Klein and Moeschberger (2003), though other kernel functions may also be used. We recommend using the smallest bandwidth for which stable variance estimates are obtainable.

With Theorem 1, a $100(1 - \alpha)\%$ confidence interval for $\hat{\theta}(t, q|\mathbf{Z})$ can be calculated by taking $\hat{\theta}(t, q|\mathbf{Z}) \pm z_{1-\alpha/2}^* \sqrt{\hat{V}(t)}$. We can also use the asymptotic results to obtain confidence

bands, as laid out in Fleming and Harrington (2005) and Cheng (1998). Specifically, we know that

$$\sup_{0 \leq t \leq \tau} \frac{\sqrt{n}|\hat{\theta}(t, q|\mathbf{Z}) - \theta(t, q|\mathbf{Z})|}{V(\tau)} \rightarrow_p \sup_{0 \leq t \leq 1} |U(t)|$$

where $U(t)$ is standard Brownian motion. We can therefore solve

$$\Pr \left(\sup_{0 \leq t \leq 1} |U(t)| \leq c \right) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp \left\{ -\frac{\pi^2(2k+1)^2}{8c^2} \right\} \approx 1 - \alpha$$

for c and compute $100(1 - \alpha)\%$ confidence bands using $\hat{\theta}(t, q|\mathbf{Z}) \pm c \sqrt{\hat{V}(\tau)}$.

Additionally, we can compute both confidence bands and intervals without appealing to asymptotic results by using either bootstrap or simulation methods. Using the bootstrap method, we find the 2.5%ile and 97.5%ile of the bootstrap sample estimates of $\hat{\theta}^*(t, q|\mathbf{Z})$ at each time point in order to get pointwise intervals (Efron and Tibshirani, 1998). To obtain bands, we first calculate the maximum deviation within each bootstrap sample across time, $\sup_{0 \leq t \leq \tau} |\hat{\theta}^*(t, q|\mathbf{Z}) - \hat{\theta}(t, q|\mathbf{Z})|$, and then find the 97.5%ile across samples, c . We can then construct bands using $\hat{\theta}(t, q|\mathbf{Z}) \pm c$.

The simulation method is another resampling method (Parzen et al., 1994). In the formulation of $\hat{\theta}(t, q|\mathbf{Z}) - \theta(t, q|\mathbf{Z})$, we replace $dM_i(u)$ with $G_i dN_i(u)$ where $G_i \sim N(0, 1)$. For each simulated sample, all G_i are randomly generated and we compute the estimated residual time quantile for that sample, $\hat{\theta}(t, q|\mathbf{Z})$. At each time point, the 2.5%ile and 97.5%ile of the deviations $\hat{\theta}(t, q|\mathbf{Z}) - \theta(t, q|\mathbf{Z})$ are calculated and added to the estimate $\hat{\theta}(t, q|\mathbf{Z})$ to get lower and upper pointwise confidence intervals, respectively. Calculating bands is the same as with the bootstrap: we find the maximum deviation within each simulated sample across time, $\sup_{0 \leq t \leq \tau} |\hat{\theta}(t, q|\mathbf{Z}) - \hat{\theta}(t, q|\mathbf{Z})|$, and then find the 97.5%ile across samples, c . We can then construct bands using $\hat{\theta}(t, q|\mathbf{Z}) \pm c$.

While being able to estimate covariate-specific residual time quantiles and their variance is useful, in some practical applications it is also important to be able to carry out comparisons between different covariate values and determine if any observed difference is statistically significant. That is, we may be interested in formally comparing residual times at different fixed time points or for different quantiles. These tasks require being able to estimate the covariance between two different residual time quantiles, and establish the asymptotic properties of their differences, as shown in the following Corollary,

Corollary 1

We assume conditions A–D (see Appendix) of Andersen and Gill (1982) hold. If, as previously developed, $\hat{\theta}(t, q|\mathbf{Z}) = \hat{\Lambda}_0^{-1} \{ \hat{\Lambda}_0(t) - \log q \exp(-\hat{\beta}^T \mathbf{Z}) \} - t$, then as $n \rightarrow \infty$, for a given $\mathbf{Z}_1$ and $\mathbf{Z}_2$ and fixed t_1, t_2, q_1 , and q_2 ,

1. $\hat{\theta}(t_1, q_1|\mathbf{Z}_1) - \hat{\theta}(t_2, q_2|\mathbf{Z}_2) \rightarrow_p \theta(t_1, q_1|\mathbf{Z}_1) - \theta(t_2, q_2|\mathbf{Z}_2);$
2. $\sqrt{n}[\{\hat{\theta}(t_1, q_1|\mathbf{Z}_1) - \hat{\theta}(t_2, q_2|\mathbf{Z}_2)\} - \{\theta(t_1, q_1|\mathbf{Z}_1) - \theta(t_2, q_2|\mathbf{Z}_2)\}] \rightarrow_d N(0, W(t)),$
where

$$W(t) = \frac{\mathbf{A}_1^T \{\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau)\}^{-1} \mathbf{A}_1 + \int_0^{\theta(t_1, q_1|\mathbf{Z}_1)} \frac{\lambda_0(t_1+v)}{s^{(0)}(\boldsymbol{\beta}, t_1+v)} dv}{\lambda_0(\theta(t_1, q_1|\mathbf{Z}_1) + t_1)^2} + \frac{\mathbf{A}_2^T \{\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau)\}^{-1} \mathbf{A}_2 + \int_0^{\theta(t_2, q_2|\mathbf{Z}_2)} \frac{\lambda_0(t_2+v)}{s^{(0)}(\boldsymbol{\beta}, t_2+v)} dv}{\lambda_0(\theta(t_2, q_2|\mathbf{Z}_2) + t_2)^2} - \frac{2 \left\{ \int_0^{\eta_{\min}-t'} \frac{\lambda_0(t'+v)}{s^{(0)}(\boldsymbol{\beta}, t'+v)} dv + \mathbf{A}_1^T \{\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau)\}^{-1} \mathbf{A}_2 \right\}}{\lambda_0(\theta(t_1, q_1|\mathbf{Z}_1) + t_1) \lambda_0(\theta(t_2, q_2|\mathbf{Z}_2) + t_2)},$$

$$\mathbf{A}_j = \left[\log q_j \times \mathbf{Z}_j \exp(-\boldsymbol{\beta}^T \mathbf{Z}_j) + \int_0^{\theta(t_j, q_j|\mathbf{Z}_j)} \frac{s^{(1)}(\boldsymbol{\beta}, t_j+v)}{s^{(0)}(\boldsymbol{\beta}, t_j+v)} \lambda_0(t_j+v) d(t_j+v) \right]^T$$

for $j = 1, 2$, $\eta_{\min} = \min\{\theta(t_1, q_1|\mathbf{Z}_1) + t_1, \theta(t_2, q_2|\mathbf{Z}_2) + t_2\}$, and $t' = \max\{t_1, t_2\}$;

3. $\hat{W}(t) \rightarrow_p W(t)$, where

$$\hat{W}(t) = \frac{\hat{\mathbf{A}}_1^T (\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}}, \tau))^{-1} \hat{\mathbf{A}}_1 + \int_0^{\hat{\theta}(t_1, q_1|\mathbf{Z}_1)} \frac{d\hat{\lambda}_0(t_1+v)}{S^{(0)}(\hat{\boldsymbol{\beta}}, t_1+v)} dv}{\hat{\lambda}_0(\hat{\theta}(t_1, q_1|\mathbf{Z}_1) + t_1)^2} + \frac{\hat{\mathbf{A}}_2^T (\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}}, \tau))^{-1} \hat{\mathbf{A}}_2 + \int_0^{\hat{\theta}(t_2, q_2|\mathbf{Z}_2)} \frac{d\hat{\lambda}_0(t_2+v)}{S^{(0)}(\hat{\boldsymbol{\beta}}, t_2+v)} dv}{\hat{\lambda}_0(\hat{\theta}(t_2, q_2|\mathbf{Z}_2) + t_2)^2} - \frac{2 \left\{ \int_0^{\hat{\eta}_{\min}-t'} \frac{d\hat{\lambda}_0(t'+v)}{S^{(0)}(\hat{\boldsymbol{\beta}}, t'+v)} dv + \hat{\mathbf{A}}_1^T (\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}}, \tau))^{-1} \hat{\mathbf{A}}_2 \right\}}{\hat{\lambda}_0(\hat{\theta}(t_1, q_1|\mathbf{Z}_1) + t_1) \lambda_0(\hat{\theta}(t_2, q_2|\mathbf{Z}_2) + t_2)},$$

$$\hat{\mathbf{A}}_j = \left[\log q_j \times \mathbf{Z}_j \exp(-\hat{\boldsymbol{\beta}}^T \mathbf{Z}_j) + \int_0^{\hat{\theta}(t_j, q_j|\mathbf{Z}_j)} \frac{S^{(1)}(\hat{\boldsymbol{\beta}}, t_j+v)}{\{S^{(0)}(\hat{\boldsymbol{\beta}}, t_j+v)\}^2} d\bar{N}(t_j+v) \right]^T$$

for $j = 1, 2$, and $\hat{\eta}_{\min} = \min\{\hat{\theta}(t_1, q_1|\mathbf{Z}_1) + t_1, \hat{\theta}(t_2, q_2|\mathbf{Z}_2) + t_2\}$.

Proof of Corollary 1 is in the Appendix. Requisite estimation of the hazard function is carried out as previously detailed.

With this corollary, calculating pointwise confidence intervals is possible. For a $100(1-\alpha)\%$ confidence interval, take $\{\hat{\theta}(t_1, q_1|\mathbf{Z}_1) - \hat{\theta}(t_2, q_2|\mathbf{Z}_2)\} \pm z_{1-\alpha/2}^* \sqrt{\hat{W}(t)}$. Methods for calculating confidence bands and other methods for calculating confidence intervals are

similar to those described following Theorem 1, with the substitution of $\{\hat{\theta}(t_1, q_1|\mathbf{Z}_1) - \hat{\theta}(t_2, q_2|\mathbf{Z}_2)\}$ for $\hat{\theta}(t, q|\mathbf{Z})$ and $\hat{W}(t)$ for $\hat{V}(t)$.

It bears mentioning that whether one uses confidence intervals or confidence bands (and, in turn, over what time period the confidence bands are calculated) should be determined by the scientific question and that different choices may yield different results. Confidence intervals calculated at different time points may have discrepant results both from each other and from the bands obtained for the entire time interval. Not all results are possible, though: while one may find a significant difference between two treatment groups at certain time points and no significant difference over the entire time interval, the reverse is not possible. If there is a significant difference between two treatment groups across the entire time interval, then there should be no points in time at which there is no significant difference between the two treatment groups based on the pointwise intervals.

3 Numerical Studies

3.1 Simulations

In order to demonstrate the asymptotic performance of our estimators $\hat{\theta}(t, q|\mathbf{Z})$ and $\hat{V}(t)$ we conducted a simulation study. Survival times were generated under the assumption of the Cox proportional hazards model with a baseline hazard of $\lambda(t|\mathbf{Z}_i) = t \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)$ with $\boldsymbol{\beta} = (1, 2)^T$ and $\mathbf{Z}_i = (Z_{i1}, Z_{i2})^T$, where $Z_{i1} \sim \text{Bernoulli}(0.5)$ and $Z_{i2} \sim \text{Uniform}(0, 1)$ for $i = 1, \dots, n$. When generating censored data, censoring times followed an exponential distribution with rate parameter 0.17 or 0.61, to allow $\sim 10\%$ or $\sim 30\%$ censoring, respectively. To estimate model-based residual time quantiles of a specific covariate, we considered median ($q = 0.5$) residual time at $t = 0.25$ and $t = 0.75$ for $z_1 = 0$ and $z_2 = 0.5$. For this and all following simulations we used a bandwidth of 1 for kernel-smoothed hazard function estimation. Calculated quantities included the bias, simulated standard error (i.e., the standard error of the MRT estimates), mean standard error (i.e., the mean of the standard error estimates), the coverage probability of nominal 95% confidence intervals, and the average 95% confidence interval length. We used 10,000 replications in our simulation.

Simulation results can be seen in Table 1. For $n = 50$, there were insufficient events past $t = 0.75$ to calculate results. Note that the true MRTs were 0.51 at $t = 0.25$ and 0.29 at $t = 0.75$. Relative to these values, bias is negligible at larger sample sizes. Simulated standard errors (SSEs) and mean standard errors (MSEs) were very close to each other for all sample sizes but $n = 50$, for which our analytical estimates (MSEs) slightly overestimated those observed in simulation (SSEs). Coverage probabilities closely matched nominal confidence levels at larger sample sizes. As mentioned in Section 2, bandwidth was chosen as the smallest value for which stable variance estimates were obtained. For this and other simulations, smaller bandwidths would occasionally result in insufficient events and therefore unrealistically large (or even infinite) variance estimates. Larger bandwidths had relatively little effect on simulation results.

These simulations were repeated with $q = 0.1$ and $q = 0.9$ to evaluate the asymptotic performance of our estimator at high and low quantiles, respectively. Performance is

somewhat worse than for $q = 0.5$, particularly at small sample sizes, but is still adequate. Detailed results can be seen in Online Resources 2 and 3.

In order to demonstrate the performance of our estimator over time as well as the various methods of constructing confidence intervals and bands, we plotted results from a single replication of simulated data. As can be seen in Figure 1, the analytical and simulation methods give similar results, with similar widths for both pointwise confidence intervals and confidence bands. The bootstrap method yields different results; in particular, the confidence bands are much wider.

The analytical method offers the advantages of direct calculation and reproducibility, but relies on the accuracy of the asymptotic result; for smaller sample sizes it may not perform well. The simulation and bootstrap methods avoid reliance on the asymptotic result at the cost of being both computationally intensive and not as reproducible. For very large sample sizes, these methods may not be feasible at all. For the remainder of this paper we will default to using the analytical method for confidence interval and band calculation, unless otherwise noted.

We also conducted simulations to evaluate the performance of $\hat{W}(t)$. Survival times were generated with a baseline hazard specified as $\lambda(t|Z_i) = t \exp(\beta^T Z_i)$. We used $\beta = 1$ and $Z_i \sim \text{Bernoulli}(0.5)$ for $i = 1, \dots, n$. Censoring times again followed an exponential distribution, but with rate parameters 0.11 and 0.38 for $\sim 10\%$ and $\sim 30\%$ censoring, respectively. We then considered the difference $\{\hat{\theta}(t_1, q_1|z_1) - \hat{\theta}(t_2, q_2|z_2)\}$ where $t_1 = 0.25$, $q_1 = 0.4$, $z_1 = 0$, $t_2 = 0.75$, $q_2 = 0.6$, and $z_2 = 1$. We calculated bias, simulated standard error, mean standard error, and the coverage probability and average length of 95% confidence intervals associated with the difference $\{\hat{\theta}(t_1, q_1|z_1) - \hat{\theta}(t_2, q_2|z_2)\}$. We used 10,000 replications in our simulation.

Results for the performance of $\hat{W}(t)$ can be seen in Table 2. Note that the true difference in MRTs is 0.91. There is small but noticeable bias for $n = 50$, as well as slight overestimation of the standard error. These issues disappear as sample size increases.

Finally, we conducted simulations to evaluate the sensitivity of our method to model misspecification. Instead of using the proportional hazards model, survival times were generated under the assumption of additive hazards with a baseline hazard of $\lambda(t|Z_i) = t + \beta^T Z_i$ with $\beta = (1, 2)^T$ and $Z_i = (Z_{i1}, Z_{i2})^T$, where $Z_{i1} \sim \text{Bernoulli}(0.5)$ and $Z_{i2} \sim \text{Uniform}(0, 1)$. Censoring times followed an exponential distribution with rate parameter 0.195 ($\sim 10\%$ censoring) or 0.72 ($\sim 30\%$ censoring).

For the purposes of simulation, we considered median ($q = 0.5$) residual time at $t = 0.25$ and $t = 0.75$. Covariate values of $z_1 = 1$ and $z_2 = 0.5$ were chosen. Importantly, we estimated median residual times using our Cox model-based estimator described above. Calculated quantities included the bias, simulated standard error, mean standard error, and the coverage probability and average length of nominal 95% confidence intervals. We used 10,000 replications in our simulation.

Results from this simulation can be seen in Table 3. The true MRTs are 0.29 at $t = 0.25$ and 0.24 at $t = 0.75$. Bias is relatively small but definitely present. Standard error estimates are surprisingly close to those observed via simulation, but coverage is still poor, particularly for $t = 0.75$. While our estimator's performance is clearly worse when applied to data that is not generated under a proportional hazards model, it does not seem so bad as to cause particular concern, at least according to the results of this simulation.

3.2 Real data examples

We present two examples of analysis on existing data sets. For the first, we use data from a clinical trial investigating the treatment of carcinoma of the oropharynx as presented in Kalbfleisch and Prentice (2002). These data include 195 patients with squamous cell carcinoma of 3 sites in the oropharynx from the 6 largest of 16 total participating institutions. Patients were randomly assigned treatment with radiation therapy alone or radiation therapy with a chemotherapeutic agent. While many other characteristics were collected, the data set we examined included only: sex, treatment, tumor grade and site, tumor (T) staging, lymph node (N) staging, overall patient condition, date of entry, living status, and survival time. Of the 195 patients included, 142 were observed to fail.

We examined differences in MRTs to death across varying T staging categories: I/II, III, or IV. Regression results from the Cox model can be seen in Table 4. Neither variable violated the proportionality assumption to a statistically significant degree. The regression results reflect what we see when plotting MRTs (Figure 2): higher T stage is associated with increased hazard of death and lower MRT. The relationship between T stage and MRT is not statistically significant, however. As can be seen in Online Resource 4, the null value of 0 is well within the confidence bands for the difference between MRTs for any two values of T stage. It may be worth noting that for certain fixed time points, the difference in MRTs for T stages IV vs I/II and for IV vs III are significantly different from 0.

While not statistically significant, our results highlight a benefit of examining MRTs: for patients with T staging of I, II, or III, MRT changes markedly over time, increasing sharply after initially decreasing slightly. For example, patients with T stage III have a MRT of 422 days (95% CI 266–578 days) after having survived 200 days and a MRT of 855 days (95% CI 350–1360 days) after having survived 450 days. This would be important and welcome news to surviving patients and their caregivers alike.

As with our simulations, in this example bandwidth was chosen as the smallest value for which stable variance estimates were obtained. This resulted in a bandwidth of 500 days. Choice of smaller bandwidths tended to result in insufficient events and therefore unrealistically large (or even infinite) variance estimates while choice of larger bandwidths had relatively little effect on estimation.

Our second example uses data collected as part of the HIVNET 012 randomized trial (Jackson et al., 2003). This trial randomly assigned HIV-1 infected pregnant women in Kampala, Uganda to either a nevirapine- or zidovudine-based treatment. Their infants were followed and tested at pre-determined intervals for HIV-1. Data were also collected on adverse events through 6–8 weeks postpartum for mothers and 18 months for babies. The

study enrolled 645 mothers: 313 assigned to nevirapine (NVP), 313 to zidovudine (AZT), and 19 to placebo. Within 18 months, 109 serious adverse events and 34 deaths were observed in the nevirapine group while 97 serious adverse events and 42 deaths were observed in the zidovudine group.

For our analysis we examined the relationship between the 10%ile of residual time to death or serious adverse event and treatment group (nevirapine versus zidovudine), maternal CD4 at pre-entry, and maternal HIV-1 RNA at pre-entry. We present results from the Cox model in Table 4. None of the variables violated the proportionality assumption to a statistically significant degree. In the multivariate model, only maternal HIV-1 RNA was significantly associated with time to death or serious adverse event, with a hazard ratio of 1.41 (95% CI 1.15–1.72) for a unit increase of $\log_{10}$ HIV-1 RNA copies/mL. A plot of 10%ile residual times for two different combinations of covariate values (a theoretical infant whose mother was treated with AZT, had a maternal CD4 of 600 cells/ μ L, and a $\log_{10}$ maternal viral load of 3.5 copies/mL and a theoretical infant whose mother was treated with NVP, had a maternal CD4 of 300 cells/ μ L, and a $\log_{10}$ maternal viral load of 5 copies/mL) can be seen in Figure 3, where the combinations were chosen to represent infants with relatively protective or relatively non-protective characteristics. For both groups, the 10%ile residual times are fairly stable, particularly at early follow-up times. For any amount of time survived event-free less than 150 days, we expect 10% of infants to die or experience a serious adverse event within the next ~ 200 days for the low-risk group and ~ 100 days for the high-risk group.

When examining the difference in 10%ile residual times between the two different combinations of covariate values we conclude that it is not significant across all time as the null value of 0 is well within the limits of the confidence bands (see Online Resource 5). This is in spite of the fact that, for the majority of fixed time points, the difference is statistically significant. The lack of overall significance is driven to a large degree by the large increase in variance at later times.

Again, bandwidth for kernel-smoothed hazard estimation was chosen as the smallest value for which stable variance estimates were obtained. This resulted in a bandwidth of 50 days. As before, choice of smaller bandwidths resulted in insufficient events and unrealistically large variance estimates while choice of larger bandwidths had relatively little effect on estimation.

4 Discussion

In this article, we presented a model-based method for estimation of residual time quantiles for censored time-to-event outcomes. As discussed earlier, in clinical practice knowing individuals' covariate-specific residual time quantiles may assist both clinicians and patients in better understanding how individually carried risk factors affect an event time such as cancer survival. The model-based estimation of residual time quantiles that we propose harnesses the popularity of the Cox proportional hazards model to provide a simple approach to estimating residual time quantiles with easy-to-compute variances.

A couple aspects of our method bear further discussion. One is the choice of bandwidth when estimating the hazard function for the variance calculation. This is particularly important for low or high quantiles, such as the 10%ile and 90%ile used in some of our simulations. Inaccurate variance estimation and in turn incorrect coverage probability may result from a poorly chosen bandwidth. Additionally, as with all asymptotic results, ours may not hold with smaller sample sizes. Sample sizes smaller than $n = 50$ or with relatively few observed events may result in poor estimates and possibly the inability to calculate estimates at all quantiles or time points.

As we explored to some degree in our simulations, the Cox proportional hazards model that we used for methods development is not the only modeling tool for residual time quantiles estimation and may not always fit the observed data. In particular, the proportionality assumption may be violated. One possible extension is to allow time-varying coefficients or time-varying covariates, such as covariate-time interaction, in the Cox model. This may be a promising direction for further methods development, as it would allow for both relaxation of the assumption of proportionality and additional applications. For example, a patient with asthma may be very interested in their median residual time to a severe respiratory event given not only that they have thus far been free of one but have also recently moved to a location with less air pollution.

Alternatives to the Cox model, such as the accelerated failure time model or the additive hazard model, could also be used for estimating residual time quantiles. Since the methods we propose in this article are general enough, we expect that our proposed methods can be applied similarly to these models, but estimation and asymptotics will need to be developed separately.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Andersen P, Gill R. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*. 1982; 10(4):1100–1120.
- Cheng K. On almost sure representation for quantiles of the product limit estimator with applications. *Sankhyā : The Indian Journal of Statistics, Series A*. 1984; 46(3):426–443.
- Cheng S, Fine J, Wei L. Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*. 1998; 54(1):219–228. [PubMed: 9544517]
- Dabrowska D, Doksum K. Estimates and confidence intervals for median and mean life in the proportional hazard model. *Biometrika*. 1987; 74(4):799–807.
- Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*. Vol. 57. CRC Press LLC; 1998.
- Epanechnikov V. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*. 1969; 14(1):153–158.
- Fleming, T.; Harrington, D. *Counting Processes and Survival Analysis*. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2005.
- Gelfand A, Kottas A. Bayesian semiparametric regression for median residual life. *Scandinavian Journal of Statistics*. 2003; 30(4):651–665.
- Jackson J, Musoke P, Fleming T, Guay L, Bagenda D, Allen M, Nakabiito C, Sherman J, Bakaki P, Owor M, Ducar C, Deseyve M, Mwatha A, Emel L, Duefield C, Mirochnick M, Fowler M,

- Mofenson L, Miotti P, Gigliotti M, Bray D, Mmiro F. Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: 18-month follow-up of the HIVNET 012 randomised trial. *The Lancet*. 2003; 362:859–868.
- Jeong J, Jung S, Costantino J. Nonparametric inference on median residual life function. *Biometrics*. 2008; 64(1):157–163. [PubMed: 17501936]
- Jung S, Jeong J, Bandos H. Regression on quantile residual life. *Biometrics*. 2009; 65(4):1203–1212. [PubMed: 19432781]
- Kalbfleisch, J.; Prentice, R. *The Statistical Analysis of Failure Time Data*. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2002.
- Kaplan E, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*. 1958; 53(282):457–481.
- Klein, J.; Moeschberger, M. *Survival Analysis: Techniques for Censored and Truncated Data*. New York, New York: Springer-Verlag; 2003.
- Ma Y, Yin G. Semiparametric median residual life model and inference. *Canadian Journal of Statistics*. 2010; 38(4):665–679.
- Parzen M, Wei L, Ying Z. A resampling method based on pivotal estimating functions. *Biometrika*. 1994; 81(2):341–350.
- Ramlau-Hansen H. Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics*. 1983; 11(2):453–466.
- Yandell B. Nonparametric inference for rates with censored survival data. *The Annals of Statistics*. 1983; 11(4):1119–1135.
- Yang S. Censored median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association*. 1999; 94(445):137–145.
- Ying Z, Jung S, Wei L. Survival analysis with median regression models. *Journal of the American Statistical Association*. 1995; 90(429):178–184.

Appendix

We restate conditions A–D of Andersen and Gill (1982) with minor modification as in Fleming and Harrington (2005):

- A.** The time τ is such that $\int_0^\tau \lambda(t) dt < \infty$.
- B.** For $S^{(j)}, j = 0, 1, 2$, there exists a neighborhood $\mathcal{B}$ of β and scalar, vector, and matrix functions $s^{(0)}, s^{(1)}$, and $s^{(2)}$ defined on $\mathcal{B} \times [0, \tau]$ such that, for $j = 0, 1, 2$,

$$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \|S^{(j)}(\beta, t) - s^{(j)}(\beta, t)\| \rightarrow 0$$

in probability as $n \rightarrow \infty$.

- C.** There exists $\delta > 0$ such that

$$\sqrt{n} \sup_{i, t} |Z_i| Y_i(t) I(\beta^T Z_i > -\delta |Z_i|) \rightarrow 0$$

in probability as $n \rightarrow \infty$.

- D.** Let $\mathcal{B}$ and $s^{(j)}, j = 0, 1, 2$ be defined as in Condition B and e and v as in Section 2. Then for all $\beta \in \mathcal{B}$ and $t \in [0, \tau]$:

$$s^{(1)}(\beta, t) = \frac{\partial}{\partial \beta} s^{(0)}(\beta, t), s^{(2)}(\beta, t) = \frac{\partial^2}{\partial \beta^2} s^{(0)}(\beta, t),$$

$s^{(0)}(\cdot, t)$, $s^{(1)}(\cdot, t)$, and $s^{(2)}(\cdot, t)$ are continuous function of $\beta \in \mathcal{B}$, uniformly in $t \in [0, \tau]$, $s^{(0)}$, $s^{(1)}$, and $s^{(2)}$ are bounded on $\mathcal{B} \times [0, \tau]$; $s^{(0)}$ is bounded away from zero on $\mathcal{B} \times [0, \tau]$, and the matrix Σ (as defined in Section 2) is positive definite.

Proof (Theorem 1)

The proof for (1) follows directly from the established convergence of the individual estimators and the application of the continuous mapping theorem. Likewise for (3), where the consistency of the kernel-estimated baseline hazard has already been established (Ramlau-Hansen, 1983; Yandell, 1983). Therefore only (2) requires a detailed proof.

We begin with two equations: one based on the true values,

$$\Lambda_0\{\theta(t, q|\mathbf{Z})+t\} = \Lambda_0(t) - \log q \times \exp(-\beta^T \mathbf{Z}),$$

and one based on the estimated values,

$$\hat{\Lambda}_0\{\hat{\theta}(t, q|\mathbf{Z})+t\} = \hat{\Lambda}_0(t) - \log q \times \exp(-\hat{\beta}^T \mathbf{Z}).$$

Taking the differences between the left- and right-hand sides of both equations yields

$$\hat{\Lambda}_0\{\hat{\theta}(t, q|\mathbf{Z})+t\} - \Lambda_0\{\theta(t, q|\mathbf{Z})+t\} = \hat{\Lambda}_0(t) - \log q \times \exp(-\hat{\beta}^T \mathbf{Z}) - \Lambda_0(t) + \log q \times \exp(-\beta^T \mathbf{Z}).$$

Examining the left-hand side first, we have

$$\begin{aligned} & \hat{\Lambda}_0\{\hat{\theta}(t, q|\mathbf{Z})+t\} - \Lambda_0\{\theta(t, q|\mathbf{Z})+t\} \\ &= \hat{\Lambda}_0\{\hat{\theta}(t, q|\mathbf{Z})+t\} - \hat{\Lambda}_0\{\theta(t, q|\mathbf{Z})+t\} \\ &+ \hat{\Lambda}_0\{\theta(t, q|\mathbf{Z})+t\} - \Lambda_0\{\theta(t, q|\mathbf{Z})+t\} \\ &\approx \hat{\lambda}_0\{\theta(t, q|\mathbf{Z})+t\}\{\hat{\theta}(t, q|\mathbf{Z}) - \theta(t, q|\mathbf{Z})\} \\ &+ \hat{\Lambda}_0\{\theta(t, q|\mathbf{Z})+t\} - \Lambda_0\{\theta(t, q|\mathbf{Z})+t\}, \end{aligned}$$

where the approximation is due to Taylor's expansion. Examining the right-hand side, we have

$$\hat{\Lambda}_0(t) - \Lambda_0(t) - \log q \times \{\exp(-\hat{\beta}^T \mathbf{Z}) - \exp(-\beta^T \mathbf{Z})\} \approx \hat{\Lambda}_0(t) - \Lambda_0(t) + \log q \times \mathbf{Z} \exp(-\hat{\beta}^T \mathbf{Z})(\hat{\beta} - \beta),$$

again using Taylor's approximation. Combining everything gives us the expression

$$\begin{aligned} & \hat{\lambda}_0\{\theta(t, q|\mathbf{Z})+t\}\{\hat{\theta}(t, q|\mathbf{Z}) - \theta(t, q|\mathbf{Z})\} + \hat{\Lambda}_0\{\theta(t, q|\mathbf{Z})+t\} \\ & - \Lambda_0\{\theta(t, q|\mathbf{Z})+t\} \approx \hat{\Lambda}_0(t) - \Lambda_0(t) + \log q \times \mathbf{Z} \exp(-\hat{\boldsymbol{\beta}}^T \mathbf{Z})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \end{aligned}$$

Rearranging yields the approximation

$$\begin{aligned} \hat{\theta}(t, q|\mathbf{Z}) - \theta(t, q|\mathbf{Z}) & \approx \frac{1}{\hat{\lambda}_0\{\theta(t, q|\mathbf{Z})+t\}} \left(\log q \times \mathbf{Z} \exp(-\hat{\boldsymbol{\beta}}^T \mathbf{Z})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right. \\ & \left. + \hat{\Lambda}_0(t) - \Lambda_0(t) - \left[\hat{\Lambda}_0\{\theta(t, q|\mathbf{Z})+t\} - \Lambda_0\{\theta(t, q|\mathbf{Z})+t\} \right] \right). \end{aligned}$$

From Fleming and Harrington (2005) we know that

$$\hat{\Lambda}_0(t) - \Lambda_0(t) \approx - \left[\int_0^t \frac{\mathbf{S}^{(1)}(\hat{\boldsymbol{\beta}}, u)}{n\{S^{(0)}(\hat{\boldsymbol{\beta}}, u)\}^2} d\bar{N}(u) \right]^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \int_0^t \frac{\sum_{i=1}^n dM_i(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}, u)},$$

so we can rewrite the overall approximation as

$$\begin{aligned} \hat{\theta}(t, q|\mathbf{Z}) - \theta(t, q|\mathbf{Z}) & \approx \frac{1}{\hat{\lambda}_0\{\theta(t, q|\mathbf{Z})+t\}} \left\{ - \int_t^{\theta(t, q|\mathbf{Z})+t} \frac{\sum_{i=1}^n dM_i(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}, u)} \right. \\ & \left. + \left[\log q \times \mathbf{Z} \exp(-\hat{\boldsymbol{\beta}}^T \mathbf{Z}) + \int_t^{\theta(t, q|\mathbf{Z})+t} \frac{\mathbf{S}^{(1)}(\hat{\boldsymbol{\beta}}, u)}{n\{S^{(0)}(\hat{\boldsymbol{\beta}}, u)\}^2} d\bar{N}(u) \right]^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right\}. \end{aligned}$$

We can also perform a substitution, setting $u = t + v$, and integrating with respect to v . This yields the approximation

$$\begin{aligned} \hat{\theta}(t, q|\mathbf{Z}) - \theta(t, q|\mathbf{Z}) & \approx \frac{1}{\hat{\lambda}_0\{\theta(t, q|\mathbf{Z})+t\}} \left\{ - \int_0^{\theta(t, q|\mathbf{Z})} \frac{\sum_{i=1}^n dM_i(t+v)}{S^{(0)}(\hat{\boldsymbol{\beta}}, t+v)} \right. \\ & \left. + \left[\log q \times \mathbf{Z} \exp(-\hat{\boldsymbol{\beta}}^T \mathbf{Z}) + \int_0^{\theta(t, q|\mathbf{Z})} \frac{\mathbf{S}^{(1)}(\hat{\boldsymbol{\beta}}, t+v)}{n\{S^{(0)}(\hat{\boldsymbol{\beta}}, t+v)\}^2} d\bar{N}(t+v) \right]^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right\}. \end{aligned}$$

Applying results from Fleming and Harrington (2005) about the asymptotic variance of $\hat{\boldsymbol{\beta}}$ and martingales, we have:

$$\text{Var}\{\hat{\theta}(t, q|\mathbf{Z}) - \theta(t, q|\mathbf{Z})\} \approx \frac{\tilde{\mathbf{A}}^T \text{Var}(\hat{\boldsymbol{\beta}}) \tilde{\mathbf{A}} + \int_0^{\theta(t, q|\mathbf{Z})} \frac{d\hat{\Lambda}_0(t+v)}{S^{(0)}(\hat{\boldsymbol{\beta}}, t+v)}}{\hat{\lambda}_0\{\theta(t, q|\mathbf{Z})+t\}^2},$$

where

$$\tilde{\mathbf{A}} = \left[\log q \times \mathbf{Z} \exp(-\hat{\boldsymbol{\beta}}^T \mathbf{Z}) + \int_0^{\theta(t,q|\mathbf{Z})} \frac{\mathbf{S}^{(1)}(\hat{\boldsymbol{\beta}}, t+v)}{n\{S^{(0)}(\hat{\boldsymbol{\beta}}, t+v)\}^2} d\bar{N}(t+v) \right]^T.$$

Based on the consistency of the estimators used in this formulation, we have the asymptotic result $\sqrt{n}(\hat{\theta}(t, q|\mathbf{Z}) - \theta(t, q|\mathbf{Z})) \rightarrow_d N(0, V(t))$, where

$$V(t) \approx \frac{1}{\lambda_0\{\theta(t, q|\mathbf{Z})+t\}^2} \left[\mathbf{A}^T \{\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau)\}^{-1} \mathbf{A} + \int_0^{\theta(t,q|\mathbf{Z})} \frac{\lambda_0(t+v)}{s^{(0)}(\boldsymbol{\beta}, v)} dv \right]$$

and

$$\mathbf{A} = \left[\log q \times \mathbf{Z} \exp(-\boldsymbol{\beta}^T \mathbf{Z}) + \int_0^{\theta(t,q|\mathbf{Z})} \frac{s^{(1)}(\boldsymbol{\beta}, v)}{s^{(0)}(\boldsymbol{\beta}, v)} \lambda_0(t+v) dv \right]^T.$$

Proof (Corollary 1)

The proof for (1) follows directly from the established convergence of the individual estimators and the application of the continuous mapping theorem. Likewise for (3), where the consistency of the kernel-estimated baseline hazard has already been established (Ramlau-Hansen, 1983; Yandell, 1983). Therefore we need only prove (2) in detail.

From the proof of Theorem 1, we have

$$\hat{\theta}(t, q|\mathbf{Z}_j) - \theta(t_j, q_j|\mathbf{Z}_j) \approx \frac{1}{\hat{\lambda}_0(\theta(t_j, q_j|\mathbf{Z}_j)+t)} \times \left\{ - \int_0^{\theta(t_j, q_j|\mathbf{Z}_j)} \frac{\sum_{i=1}^n dM_j(t_j+v)}{S^{(0)}(\hat{\boldsymbol{\beta}}, t_j+v)} + \tilde{\mathbf{A}}_j^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right\}$$

where

$$\tilde{\mathbf{A}}_j = \left[\log q_j \times \mathbf{Z}_j \exp(-\hat{\boldsymbol{\beta}}^T \mathbf{Z}_j) + \int_0^{\theta(t_j, q_j|\mathbf{Z}_j)} \frac{\mathbf{S}^{(1)}(\hat{\boldsymbol{\beta}}, t_j+v)}{n\{S^{(0)}(\hat{\boldsymbol{\beta}}, t_j+v)\}^2} d\bar{N}(t_j+v) \right]^T.$$

In order to calculate $\text{Var}(\hat{\theta}(t_1, q_1|\mathbf{Z}_1) - \hat{\theta}(t_2, q_2|\mathbf{Z}_2))$, we note that

$$\begin{aligned} & \text{Var}(\hat{\theta}(t_1, q_1|\mathbf{Z}_1) - \hat{\theta}(t_2, q_2|\mathbf{Z}_2)) \\ &= \text{Var} \left[\{\hat{\theta}(t_1, q_1|\mathbf{Z}_1) - \theta(t_1, q_1|\mathbf{Z}_1)\} - \{\hat{\theta}(t_2, q_2|\mathbf{Z}_2) - \theta(t_2, q_2|\mathbf{Z}_2)\} \right] \\ &= \text{Var}\{\hat{\theta}(t_1, q_1|\mathbf{Z}_1) - \theta(t_1, q_1|\mathbf{Z}_1)\} + \text{Var}\{\hat{\theta}(t_2, q_2|\mathbf{Z}_2) - \theta(t_2, q_2|\mathbf{Z}_2)\} \\ &\quad - 2 \times \text{Cov}\{\hat{\theta}(t_1, q_1|\mathbf{Z}_1) - \theta(t_1, q_1|\mathbf{Z}_1), \hat{\theta}(t_2, q_2|\mathbf{Z}_2) - \theta(t_2, q_2|\mathbf{Z}_2)\}. \end{aligned}$$

Now

$$\begin{aligned}
& \text{Cov}\{\hat{\theta}(t_1, q_1 | \mathbf{Z}_1) - \theta(t_1, q_1 | \mathbf{Z}_1), \hat{\theta}(t_2, q_2 | \mathbf{Z}_2) - \theta(t_2, q_2 | \mathbf{Z}_2)\} \\
&= \frac{1}{\hat{\lambda}_0(\theta(t_1, q_1 | \mathbf{Z}_1) + t_1) \hat{\lambda}_0(\theta(t_2, q_2 | \mathbf{Z}_2) + t_2)} \\
&\quad \left[\text{Cov} \left\{ \int_0^{\theta(t_1, q_1 | \mathbf{Z}_1)} \frac{\sum_{i=1}^n dM_i(t_1 + v)}{S^{(0)}(\hat{\beta}, t_1 + v)}, \int_0^{\theta(t_2, q_2 | \mathbf{Z}_2)} \frac{\sum_{j=1}^n dM_j(t_2 + v)}{S^{(0)}(\hat{\beta}, t_2 + v)} \right\} \right. \\
&\quad - \text{Cov} \left\{ \int_0^{\theta(t_1, q_1 | \mathbf{Z}_1)} \frac{\sum_{i=1}^n dM_i(t_1 + v)}{S^{(0)}(\hat{\beta}, t_1 + v)}, \tilde{\mathbf{A}}_2^T (\hat{\beta} - \beta) \right\} \\
&\quad - \text{Cov} \left\{ \int_0^{\theta(t_2, q_2 | \mathbf{Z}_2)} \frac{\sum_{j=1}^n dM_j(t_2 + v)}{S^{(0)}(\hat{\beta}, t_2 + v)}, \tilde{\mathbf{A}}_1^T (\hat{\beta} - \beta) \right\} \\
&\quad \left. + \text{Cov} \left\{ \tilde{\mathbf{A}}_1^T (\hat{\beta} - \beta), \tilde{\mathbf{A}}_2^T (\hat{\beta} - \beta) \right\} \right] \\
&= \frac{1}{\hat{\lambda}_0(\theta(t_1, q_1 | \mathbf{Z}_1) + t_1) \hat{\lambda}_0(\theta(t_2, q_2 | \mathbf{Z}_2) + t_2)} \\
&\quad \left[\text{Cov} \left\{ \sum_{i=1}^n \int_0^{\eta_{\min} - t'} \frac{dM_i(t' + v)}{S^{(0)}(\hat{\beta}, t' + v)}, \sum_{j=1}^n \int_0^{\eta_{\min} - t'} \frac{dM_j(t' + v)}{S^{(0)}(\hat{\beta}, t' + v)} \right\} \right. \\
&\quad \left. - 0 - 0 + \tilde{\mathbf{A}}_1^T \text{Var}(\hat{\beta} - \beta) \tilde{\mathbf{A}}_2 \right] \\
&= \frac{1}{\hat{\lambda}_0(\theta(t_1, q_1 | \mathbf{Z}_1) + t_1) \hat{\lambda}_0(\theta(t_2, q_2 | \mathbf{Z}_2) + t_2)} \\
&\quad \left[\sum_{i=j}^n \text{Cov} \left\{ \int_0^{\eta_{\min} - t'} \frac{dM_i(t' + v)}{S^{(0)}(\hat{\beta}, t' + v)}, \int_0^{\eta_{\min} - t'} \frac{dM_j(t' + v)}{S^{(0)}(\hat{\beta}, t' + v)} \right\} \right. \\
&\quad \left. + \tilde{\mathbf{A}}_1^T \text{Var}(\hat{\beta} - \beta) \tilde{\mathbf{A}}_2 \right] \\
&= \frac{1}{\hat{\lambda}_0(\theta(t_1, q_1 | \mathbf{Z}_1) + t_1) \hat{\lambda}_0(\theta(t_2, q_2 | \mathbf{Z}_2) + t_2)} \\
&\quad \left[\text{Var} \left\{ \int_0^{\eta_{\min} - t'} \frac{\sum_{i=1}^n dM_i(t' + v)}{S^{(0)}(\hat{\beta}, t' + v)} \right\} + \tilde{\mathbf{A}}_1^T \text{Var}(\hat{\beta} - \beta) \tilde{\mathbf{A}}_2 \right],
\end{aligned}$$

where $\eta_{\min} = \min\{\theta(t_1, q_1 | \mathbf{Z}_1) + t_1, \theta(t_2, q_2 | \mathbf{Z}_2) + t_2\}$ and $t' = \max\{t_1, t_2\}$.

So, combining the above with results from Theorem 1 and taking into account the consistency of the estimators used in this formulation, we have the asymptotic result

$$\sqrt{n}[\{\hat{\theta}(t_1, q_1 | \mathbf{Z}_1) - \hat{\theta}(t_2, q_2 | \mathbf{Z}_2)\} - \{\theta(t_1, q_1 | \mathbf{Z}_1) - \theta(t_2, q_2 | \mathbf{Z}_2)\}] \rightarrow_d N(0, W(t)),$$

where

$$\begin{aligned}
W(t) = & \frac{\mathbf{A}_1^T \{ \boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau) \}^{-1} \mathbf{A}_1 + \int_0^{\theta(t_1, q_1 | \mathbf{Z}_1)} \frac{\lambda_0(t_1 + v)}{s^{(0)}(\boldsymbol{\beta}, t_1 + v)} dv}{\lambda_0(\theta(t_1, q_1 | \mathbf{Z}_1) + t_1)^2} \\
& + \frac{\mathbf{A}_2^T \{ \boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau) \}^{-1} \mathbf{A}_2 + \int_0^{\theta(t_2, q_2 | \mathbf{Z}_2)} \frac{\lambda_0(t_2 + v)}{s^{(0)}(\boldsymbol{\beta}, t_2 + v)} dv}{\lambda_0(\theta(t_2, q_2 | \mathbf{Z}_2) + t_2)^2} \\
& - \frac{2 \left\{ \int_0^{\eta_{\min} - t'} \frac{\lambda_0(t' + v)}{s^{(0)}(\boldsymbol{\beta}, t' + v)} dv + \mathbf{A}_1^T \{ \boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau) \}^{-1} \mathbf{A}_2 \right\}}{\lambda_0(\theta(t_1, q_1 | \mathbf{Z}_1) + t_1) \lambda_0(\theta(t_2, q_2 | \mathbf{Z}_2) + t_2)},
\end{aligned}$$

$$\mathbf{A}_j = \left[\log q_j \times \mathbf{Z}_j \exp(-\boldsymbol{\beta}^T \mathbf{Z}_j) + \int_0^{\theta(t_j, q_j | \mathbf{Z}_j)} \frac{s^{(1)}(\boldsymbol{\beta}, t_j + v)}{s^{(0)}(\boldsymbol{\beta}, t_j + v)} \lambda_0(t_j + v) d(t_j + v) \right]^T.$$

Fig. 1.

Plots of MRT for simulated data with various methods for calculating confidence intervals and bands. The black solid curve represents estimated MRT, the gray solid curve represents true MRT, the dashed curve represents the 95% confidence interval, and the dotted curve represent the 95% confidence band.

Fig. 2.

Plots of MRTs for carcinoma data with pointwise confidence intervals and confidence bands. The solid curves represent estimated MRT, the dashed curves represent the pointwise 95% confidence intervals, and the dotted curves represent the 95% confidence bands.

Fig. 3.

Plots of 10%ile residual times for HIVNET 012 data with pointwise confidence intervals.

The solid curves represent estimated MRT, the dashed curves represent the pointwise 95% confidence intervals, and the dotted curves represent the 95% confidence bands.

Table 1

Simulation results for asymptotic performance of MRT estimator

Time	Sample Size	% Censoring	Bias	SSE	MSE	95% CP	95% CI Length
$t = 0.25$	$n = 50$	0	0.002	0.093	0.093	0.93	0.36
		10	0.001	0.097	0.100	0.93	0.39
		30	0.005	0.112	0.123	0.93	0.48
	$n = 200$	0	0.000	0.046	0.046	0.94	0.18
		10	0.001	0.048	0.048	0.94	0.19
		30	0.001	0.055	0.055	0.94	0.22
	$n = 500$	0	0.000	0.029	0.029	0.95	0.11
		10	0.000	0.030	0.030	0.95	0.12
		30	0.000	0.035	0.035	0.95	0.14
$t = 0.75$	$n = 50$	0	-0.003	0.093	0.102	0.95	0.40
		10	-0.003	0.100	0.116	0.95	0.45
		30	-	-	-	-	-
	$n = 200$	0	-0.002	0.047	0.048	0.95	0.19
		10	-0.001	0.051	0.053	0.95	0.21
		30	0.001	0.062	0.067	0.95	0.26
	$n = 500$	0	0.000	0.030	0.030	0.95	0.12
		10	0.001	0.033	0.033	0.95	0.13
		30	0.000	0.040	0.041	0.94	0.16

Table 2
Simulation results for asymptotic performance of difference between MRT estimators

Sample Size	% Censoring	Bias	SSE	MSE	95% CP	95% CI Length
$n = 50$	0	-0.008	0.173	0.183	0.94	0.72
	10	-0.010	0.180	0.196	0.94	0.77
	30	-0.011	0.211	0.235	0.95	0.92
$n = 200$	0	-0.002	0.090	0.091	0.95	0.36
	10	-0.003	0.095	0.096	0.95	0.38
	30	-0.002	0.109	0.110	0.94	0.43
$n = 500$	0	0.000	0.060	0.057	0.95	0.22
	10	-0.001	0.061	0.061	0.94	0.24
	30	0.001	0.069	0.069	0.95	0.27

Table 3
Simulation results for asymptotic performance of MRT estimator using additive hazard data

Time	Sample Size	% Censoring	Bias	SSE	MSE	95% CP	95% CI Length
$t = 0.25$	$n = 50$	0	0.013	0.089	0.089	0.93	0.35
		10	0.013	0.091	0.093	0.94	0.36
		30	0.014	0.102	0.107	0.93	0.42
	$n = 200$	0	0.011	0.043	0.045	0.95	0.18
		10	0.012	0.045	0.047	0.95	0.18
		30	0.010	0.050	0.052	0.95	0.20
	$n = 500$	0	0.011	0.027	0.028	0.94	0.11
		10	0.011	0.029	0.030	0.94	0.12
		30	0.009	0.032	0.033	0.94	0.13
$t = 0.75$	$n = 50$	0	-0.011	0.103	0.101	0.92	0.40
		10	-	-	-	-	-
		30	-	-	-	-	-
	$n = 200$	0	-0.011	0.053	0.050	0.92	0.20
		10	-0.012	0.057	0.054	0.92	0.21
		30	-0.011	0.071	0.069	0.93	0.27
	$n = 500$	0	-0.012	0.033	0.032	0.92	0.12
		10	-0.012	0.036	0.034	0.92	0.13
		30	-0.014	0.045	0.043	0.92	0.17

Table 4

Cox model results

Data Set	Variable	Hazard Ratio	95% Confidence Interval	P value
Carcinoma	T stage I/II	1.00	—	0.007*
	T stage III	1.24	0.77 to 2.02	
	T stage IV	2.01	1.22 to 3.31	
HIVNET 012	Nevirapine	1.00	—	—
	Zidovudine	0.92	0.70 to 1.21	0.550
	Maternal CD4 [†]	0.99	0.93 to 1.05	0.670
	Maternal HIV-1 RNA [‡]	1.41	1.15 to 1.72	0.001

* From likelihood ratio test of complete versus null model.

[†] Per 100 cells/ μ L at pre-entry.

[‡] Per unit increase of log₁₀ HIV-1 RNA copies/mL at pre-entry.