Harvard University Harvard University Biostatistics Working Paper Series

Year 2008

Paper 92

Calibrating Parametric Subject-specific Risk Estimation

Tianxi Cai^{*} Lu Tian[†] Hajime Uno[‡]

Scott D. Solomon**

L. J. Wei^{††}

*Harvard University, tcai@hsph.harvard.edu

[†]Northwestern University, lutian@northwestern.edu

[‡]Dana Farber Cancer Institute, huno@hsph.harvard.edu

**Brigham & Women's Hospital

^{††}Harvard University, wei@hsph.harvard.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

http://biostats.bepress.com/harvardbiostat/paper92

Copyright ©2008 by the authors.

CALIBRATING PARAMETRIC SUBJECT-SPECIFIC RISK ESTIMATION

T. CAI¹, L. TIAN², HAJIME UNO³, SCOTT D. SOLOMON⁴, AND L.J. WEI¹

¹Department of Biostatistics, Harvard University, Boston, MA 02115, USA

²Department of Preventive Medicine, Northwestern University, Chicago, IL 60611, U.S.A

³Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Boston, Massachusetts 02115, U.S.A.

⁴Cardiovascular Division, Brigham & Women's Hospital, Boston, Massachusetts 02115, U.S.A.

SUMMARY

For modern evidence-based medicine, decisions on disease prevention or management strategies for individuals in a population are often guided by a risk index system. For each individual, the system provides an estimate using the subject-specific "baseline" information for the risk of experiencing a future disease-related clinical event. Such a risk scoring scheme is usually derived from an overly simplified parametric model. To validate a model-based procedure, one may perform a standard *global* evaluation via, for instance, the ROC analysis. In this paper, we propose a method to calibrate the risk index system at a subject level. Specifically, a consistent estimation procedure is derived for the t-year mortality rate of individuals who have the same estimated parametric risk score. Furthermore, the corresponding pointwise and simultaneous confidence interval estimators are also provided over a range of risk scores of interest. The proposals are illustrated with a data set from a large clinical trial for evaluating various therapies for treating high risk patients after myocardial infarction.

Keywords: Cardiovascular diseases; Cox's model; Nonparametric functional estimation; Risk index; ROC analysis; Survival Analysis.



1. INTRODUCTION

The choice of disease prevention or management strategy for an individual from a population of interest is often made based on her/his risk of experiencing a certain clinical event during a specific follow-up time period. Subjects classified as high risk may be recommended for intensive prevention or therapy. The risk is usually estimated via a parametric or semi-parametric regression model using 'baseline" risk factors/markers. For example, recently Eagle et al. (2004) created a simple bedside risk index system to predict 6-month post-discharge mortality for patients hospitalized due to acute coronary syndrome (ACS). To estimate each individual risk, a Cox's proportional hazards model (Cox, 1972) was utilized to fit the data from the Global Registry of Acute Coronary Events (GRACE) involving 17142 patients. The final model for prediction, consisting of nine covariates, was developed using information on post-discharge follow-up mortality and all potential predictors taken at admission as well as during hospitalization. The 6-month post-discharge mortality rate is then estimated with the fitted model for each future individual patient.

An empirical model, such as a Cox model, is merely an approximation to the true model. It is crucial to validate or evaluate the performance of a *working* model for appropriate clinical decision making. In addition to checking the fitted model with routine goodness of fit techniques, the modelbased scoring system is often evaluated or validated *globally* based on specific summary measures, for example, from the standard receiver operating characteristic curve (ROC) (Pepe et al., 2004; Wang et al., 2006; Ware, 2006; Cook, 2007; Uno et al., 2007). For the aforementioned GRACE study, the investigators used an independent cohort of 7638 patients to validate their system with an overall c-statistic. However, even if the risk index system is acceptable using a global average measure, for the group of patients with the same model-based risk score, the corresponding parametric risk estimator may not estimate its true mortality rate well. Consequently clinicians may be misguided in choosing inappropriate therapies for treating these individuals.

In this article, we propose a procedure for calibrating the parametric risk estimates. Specifically,

under a general survival analysis setting, a consistent estimation procedure is proposed for the tyear mortality rate of patients grouped by a parametric risk index system. Furthermore, we provide pointwise and simultaneous confidence intervals for such risks over a range of estimated risk scores. These interval estimates quantify the precision of our consistent point estimator. Moreover, the upper or lower bounds of the intervals provide valuable information for the complex cost-benefit decision making. Note that our proposal is also useful for calibrating a risk index system when applied to populations different from the study population. For example, the Framingham Risk Score (Anderson et al., 1991) for predicting the risk of Coronary Heart Disease (CHD) was developed based on a US population. Its applicability to other populations has been extensively investigated in recent years (D'Agostino et al., 2001; Brindle et al., 2003; Zhang et al., 2005). For a new population, our procedure can be used to assess the true risk of CHD for subjects with any given Framingham Risk Score.

We illustrate the new proposal with a data set from the "Valsartan in Acute Myocardial Infarction" (VALIANT) study (Pfeffer et al., 2003). This large clinical study was conducted to evaluate the effect of angiotensin-converting-enzyme (ACE) inhibitors on overall mortality among patients with myocardial infarction complicated by left ventricular systolic dysfunction and/or heart failure. The trial was designed to compare three treatments, ARB valsartan, ACE inhibitor captopril and a combination of these two drugs, for treating high risk patients after myocardial infarction with respect to mortality and morbidity. The study was conducted from 1999 to 2003 with a total of 14703 patients randomly, equally assigned to the three groups. The median follow-up time is 24.7 months after randomization. At the end of the studies, 979, 958 and 941 patients died in the three treatment groups, respectively. There are no significant differences among three groups with respect to the overall mortality. On the other hand, the rich VALIANT data base provides valuable information about preventing future high risk patients from developing cardiovascular-related diseases after MI with certain complications. In this paper, for simplicity, we used eleven baseline

covariates from each study subject considered by Solomon et al. (2005) in our analysis. These eleven predictors are potentially the most significant ones among the risk factors/markers identified by Anavekar et al. (2004) for the overall mortality of VALIANT patients. We fitted the survival data with an additive Cox model using these eleven covariates to obtain a parametric risk score for each individual. We then estimated the t-year mortality rate consistently for subjects with a given parametric score value to calibrate the model-based risk assessment. Lastly, we constructed simultaneous and pointwise interval estimates for making inferences about such mortality rates over a range of parametric risk scores.

2. CONSISTENT ESTIMATOR FOR MEAN RISK OF SUBJECTS WITH THE SAME MODEL-BASED RISK SCORE

Consider a subject randomly drawn from the study population. Let \tilde{T} be its time to a specific event and U be the corresponding set of "baseline" risk factors/markers. Assume that \tilde{T} has a continuous distribution given U. Also, let $\mathcal{P}(U) = \operatorname{pr}(\tilde{T} < t_0 \mid U)$, the risk of this individual experiencing the event by a specific time point t_0 conditional on U. The event time \tilde{T} may be censored by a random variable C which is assumed to be independent of \tilde{T} and U. For \tilde{T} , one can only observe $T = \min(\tilde{T}, C)$ and $\Delta = I(\tilde{T} \leq C)$, where $I(\cdot)$ is the indicator function. Our data $\{(T_i, \Delta_i, U_i), i = 1, \dots, n\}$ consist of n independent copies of (T, Δ, U) .

For a given U = u, $\mathcal{P}(u)$ may be consistently estimated nonparametrically. In practice, however, such a nonparametric functional estimate does not behave well when the dimension of U is greater than one. A standard, feasible way to reduce the dimension of u is to approximate $\mathcal{P}(u)$ with a *working* parametric or semi-parametric model such as the Cox proportional hazards model:

$$\mathcal{P}(u) = g\{\log \Lambda(t_0) + \beta' x\},\tag{2.1}$$

where $g(s) = 1 - e^{-e^s}$, $\Lambda(\cdot)$ is the *working* baseline cumulative hazard function for \tilde{T} , x, a $p \times 1$ vector, is a function of u, and β is an unknown vector of regression parameters.

To obtain an estimate for $\mathcal{P}(u)$ via (2.1), one may employ the maximum partial likelihood estimator $\hat{\beta}$ for β with all mortality information from the data collected during the study time period $[0, t_0]$. That is, $\hat{\beta}$ is the maximizer of the log partial likelihood function,

$$\sum_{i=1}^{n} \int_{0}^{t_{0}} \left[\beta' X_{i} - \log \left\{ \sum_{j=1}^{n} Y_{j}(t) e^{\beta' X_{j}} \right\} \right] dN_{i}(t),$$
(2.2)

where $N_i(t) = I(T_i \leq t)\Delta_i$ and $Y_i(t) = I(T_i \geq t)$, for $i = 1, \dots, n$. When the model (2.1) is correctly specified, $\hat{\beta}$ consistently estimates the true value of β . In Appendix A, we show that under a rather mild condition, $\hat{\beta}$ converges to a finite constant β_0 , as $n \to \infty$, even when the model (2.1) is misspecified. This large sample stabilization property is critical for establishing a risk score system.

Next, we estimate the function $\Lambda(t)$ in (2.1) with the standard Breslow's estimator (Kalbfleisch & Prentice, 2002),

$$\hat{\Lambda}(t) = \sum_{i=1}^{n} \int_{0}^{t} \frac{dN_{i}(s)}{\sum_{j=1}^{n} Y_{j}(s) e^{\hat{\beta}' X_{j}}},$$

which is a step function that only jumps at observed failure time points. It follows that a model based estimate of $\mathcal{P}(u)$ is $\hat{\mathcal{P}}(u) = g(\hat{\gamma}'\vec{x})$, where $\hat{\gamma} = (\log\{\hat{\Lambda}(t_0)\}, \hat{\beta}')'$ and $\vec{x} = (1, x')'$. Again, in Appendix A, we show that under the working model (2.1), $\hat{\mathcal{P}}(u)$ converges to a deterministic function $\bar{\mathcal{P}}(u)$ in probability, as $n \to \infty$. When (2.1) is correctly specified, $\hat{\mathcal{P}}(u)$ is consistent for $\mathcal{P}(u)$.

Although most likely (2.1) is not the true model, the parametric risk scores $\{\hat{\mathcal{P}}(u)\}$ or $\{\hat{\gamma}'\vec{x}\}$ can be used as a risk index system for grouping future subjects in a population similar to the study population. Now, consider a future subject whose $(\tilde{T}, U) = (\tilde{T}^0, U^0)$ and $\hat{\mathcal{P}}(U^0) = v$. Then it is not clear that v is a reasonable estimate for the risk of this individual experiencing the event before time t_0 . Therefore, to calibrate the subject-level model-based risk estimate, one needs a consistent estimate for

$$\tau(v; t_0) = \operatorname{pr}(\tilde{T}^0 < t_0 \mid \hat{\mathcal{P}}(U^0) = v),$$
(2.3)

the mortality rate among subjects with risk score $\hat{\mathcal{P}}(U^0) = v$. Here, the probability is with respect

5

to T^0, U^0 and $\{(T_i, U_i, \Delta_i), i = 1, \cdots, n\}.$

To this end, let $\Lambda_v(t) = -\log\{1 - \tau(v;t)\}, 0 \le t \le t_0$, the "cumulative hazard function" for future subjects whose estimated risk index score is v. First, one may use a nonparametric kernel Nelson-Aalen estimator smoothed over v for $\Lambda_v(t)$ with the triplets $\{(T_i, \Delta_i, \hat{\mathcal{P}}(U_i)), i = 1, ..., n\}$. Specifically, as for the one-sample estimation of the cumulative hazard function in standard survival analysis, we consider the class of potential estimators which are step functions over t and only jump at the observed failure time points with jump sizes $d\Lambda_v(t) = \Lambda_v(t) - \Lambda_v(t-)$. Then a nonparametric functional estimator for $d\Lambda_v(t)$ can be obtained by minimizing

$$\sum_{i=1}^{n} K_h(\hat{D}_{vi}) \{ dN_i(t) - Y_i(t) d\Lambda_v(t) \}^2,$$
(2.4)

where $dN_i(t) = N_i(t) - N_i(t-)$, $\hat{D}_{vi} = \psi\{\hat{\mathcal{P}}(U_i)\} - \psi(v)$, $K(\cdot)$ is a smooth symmetric density function, $K_h(x) = K(x/h)/h$, h is a bandwidth such that $h \to 0$ and $nh^2 \to \infty$ as $n \to \infty$, and $\psi(\cdot) : (0,1) \to (-\infty,\infty)$ is a known increasing smooth transformation function (Wand et al., 1991; Park et al., 1997). This local constant estimator essentially assumes that for v' in a small neighborhood of v, $d\Lambda_{v'}(t) \approx d\Lambda_v(t)$. The resulting minimizer of (2.4) is

$$d\tilde{\Lambda}_v(t) = \frac{\sum_{i=1}^n K_h(\hat{D}_{vi}) dN_i(t)}{\sum_{i=1}^n K_h(\hat{D}_{vi}) Y_i(t)}$$

The corresponding estimator for $\Lambda_v(t_0)$ is

$$\tilde{\Lambda}_{v}(t_{0}) = \int_{0}^{t_{0}} \frac{\sum_{i=1}^{n} K_{h}(\hat{D}_{vi}) dN_{i}(t)}{\sum_{i=1}^{n} K_{h}(\hat{D}_{vi}) Y_{i}(t)}.$$
(2.5)

When $\hat{\mathcal{P}}(U_i)$ is subject-specific, not a function of the data, the limiting distribution of such a nonparametric estimator with a *fixed* v has been derived by Dabrowska (1989) and Du & Akritas (2002). Here, we are interested in the behavior of the estimator (2.5) over a set of risk scores v at a specific time point t_0 .

The above kernel estimator can potentially be improved by considering a local linear, not constant, approximation to $d\Lambda_v(t)$ around v (Fan & Gijbels, 1996; Li & Doss, 1995). That is, we

replace $d\Lambda_v(t)$ in (2.4) by a local linear function $a + b\{\psi(v') - \psi(v)\}$ with unknown intercept and slope parameters a and b, where v' is in a "neighborhood" of v. The resulting estimator $d\Lambda_v(t)$ is the intercept "a" of the vector (a, b)' which minimizes

$$\sum_{i=1}^{n} K_h(\hat{D}_{vi}) \left\{ dN_i(t) - Y_i(t)(a + \hat{D}_{vi}b) \right\}^2.$$
(2.6)

The corresponding estimator $\hat{\Lambda}_v(t)$ for $\Lambda_v(t)$ is the sum of $d\hat{\Lambda}_v(\cdot)$ over all distinct observed death times by t. The resulting estimator for $\tau(v; t_0)$ is $\hat{\tau}(v; t_0) = 1 - \exp\{-\hat{\Lambda}_v(t_0)\}$.

In Appendix B, we show that under mild regularity conditions, if $h = O(n^{-\nu})$ with $1/5 < \nu < 1/2$, $\hat{\tau}(v; t_0)$ is consistent for $\tau(v; t_0)$, uniformly in $v \in \mathcal{J} = [\psi^{-1}(\mathfrak{q}_l + h), \psi^{-1}(\mathfrak{q}_r - h)]$, where $(\mathfrak{q}_l, \mathfrak{q}_r)$ is a subset contained in the support of $\psi\{\bar{\mathcal{P}}(U)\}$.

3. POINTWISE AND SIMULTANEOUS INTERVAL ESTIMATION PROCEDURES FOR $\tau(v; t_0)$ OVER RISK SCORE v

For any fixed $v \in \mathcal{J}$, we show in Appendix B that for large n and $h = O(n^{-\nu})$ with $1/5 < \nu < 1/2$, the distribution of

$$(nh)^{\frac{1}{2}} \left\{ \hat{\Lambda}_{v}(t_{0}) - \Lambda_{v}(t_{0}) \right\}$$
 (3.1)

can be approximated well by the conditional distribution of a mean-zero normal

$$n^{-\frac{1}{2}}h^{\frac{1}{2}}\sum_{i=1}^{n}K_{h}(\hat{D}_{vi})\hat{\mathcal{V}}_{i}(v;t_{0})V_{i}+(nh)^{\frac{1}{2}}\left\{\hat{\Lambda}_{v}(t_{0},\hat{\gamma}^{*})-\hat{\Lambda}_{v}(t_{0})\right\},$$
(3.2)

given the data, where $\{V_i, i = 1, \dots, n\}$ are standard normal random variables generated independent of the data,

$$\hat{\mathcal{V}}_{i}(v;t_{0}) = \int_{0}^{t_{0}} \frac{dN_{i}(s) - Y_{i}(s)d\hat{\Lambda}_{v}(s)}{n^{-1}\sum_{j=1}^{n}K_{h}(\hat{D}_{vj})Y_{j}(s)},$$

 $\hat{\Lambda}_v(t_0, \gamma)$ is given in (B·3) and $\hat{\gamma}^*$ is given in (B·6). Here, $\hat{\Lambda}_v(t_0, \hat{\gamma}^*)$ is a random *perturbed* version of the observed $\hat{\Lambda}_v(t_0)$ with the same perturbation random sample $\{V_i, i = 1, \dots, n\}$ which accounts for the extra variability in $\hat{\gamma}$. Although asymptotically one only needs the first term of (3.2) due to

the fact that $\hat{\gamma}$ from the working Cox model is of order root-n, the inclusion of the second term is likely to improve the approximation to the distribution of (3.1) in finite sample. This perturbation method is similar to the so-called wild bootstrapping (Wu, 1986; Härdle, 1990; Mammen, 1992) and has been utilized successfully for a number of interesting problems in survival analysis (Jin et al., 2001; Park & Wei, 2003; Cai et al., 2005). Note that the distribution of (3.2) can be easily approximated by generating a large number, say M, of realizations of $\{V_i, i = 1, \dots, n\}$.

With the above large sample approximation, for any $v \in \mathcal{J}$, one may obtain a variance estimator of (3.1), $\hat{\sigma}_v^2(t_0)$, based on the empirical variance of M realizations from (3.2). For any given $\alpha \in$ (0, 1), a $100(1 - \alpha)\%$ confidence interval for $\Lambda_v(t_0)$ can then be obtained as

$$\hat{\Lambda}_v(t_0) \pm (nh)^{-1/2} c \hat{\sigma}_v(t_0),$$

where c is the $100(1 - \alpha/2)$ th percentile of the standard normal. Subsequently, the corresponding pointwise $100(1 - \alpha)\%$ confidence interval for $\tau(v; t_0)$, the mortality rate with score v, is

$$1 - \exp\left\{-\hat{\Lambda}_{v}(t_{0}) \pm (nh)^{-1/2}c\hat{\sigma}_{v}(t_{0})\right\}.$$
(3.3)

To make inference about the mortality rate over a range of v, one may construct simultaneous confidence intervals for $\{\tau(v; t_0), v \in \mathcal{J}\}$ by considering a sup-type statistic

$$W = \sup_{v \in \mathcal{J}} \left| (nh)^{\frac{1}{2}} \frac{\hat{\Lambda}_{v}(t_{0}) - \Lambda_{v}(t_{0})}{\hat{\sigma}_{v}(t_{0})} \right|.$$
(3.4)

However, as a process in v, the distribution of $(nh)^{\frac{1}{2}}\{\hat{\Lambda}_v(t_0) - \Lambda_v(t_0)\}$ does not converge to a proper distribution, as $n \to \infty$. Therefore, we cannot use the standard large sample theory for stochastic processes to obtain a finite sample approximation to the distribution of W. On the other hand, by the strong approximation arguments and extreme value limit theorem (Bickel & Rosenblatt, 1973), in Appendix B, we show that a standardized version of W converges in distribution to a proper random variable. In practice, for large n, one can approximate the distribution of W by W^* , the sup of the absolute value of (3.2) divided by $\hat{\sigma}_v(t_0)$, with (3.2) perturbed by the same set of perturbation

variables $\{V_i, i = 1, \dots, n\}$ for all $v \in \mathcal{J}$. It follows that the $100(1 - \alpha)\%$ simultaneous confidence interval for $\tau(v; t_0)$ is

$$1 - \exp\left\{-\hat{\Lambda}_{v}(t_{0}) \pm (nh)^{-1/2} d \ \hat{\sigma}_{v}(t_{0})\right\},\tag{3.5}$$

where the cutoff point d is chosen such that $pr(W^* < d) \ge 1 - \alpha$.

As for any nonparametric functional estimation problem, the choice of the smoothing parameter h for $\hat{\tau}(v; t_0)$ is crucial for making inferences about $\tau(v; t_0)$. To incorporate censoring, we propose to obtain an "optimal" h by minimizing mean integrated squared martingale residuals over time interval $(0, t_0)$ through the standard K-fold cross validation procedure. Such a procedure has been successfully used for bandwidth selection in Tian et al. (2005). Specifically, we randomly split the data into K disjoint subsets of about equal sizes denoted by $\{\mathcal{J}_k, k = 1, \dots, K\}$. For each k, we use all observations which are not in \mathcal{J}_k to obtain $\Lambda_v(t)$ with a given h. Let the resulting estimators be denoted by $\hat{\Lambda}_{v(k)}(t)$. We then use the observations from \mathcal{I}_k to calculate the sum of integrated squared martingale residuals

$$\int_{0}^{t_{0}} \sum_{j \in \mathcal{I}_{k}} \left\{ N_{j}(t) - \int_{0}^{t} Y_{j}(s) d\hat{\Lambda}_{\hat{v}_{j}(k)}(s) \right\}^{2} d\left\{ \sum_{j' \in \mathcal{I}_{k}} N_{j'}(t) \right\},$$
(3.6)

where $\hat{v}_j = \hat{\mathcal{P}}(u_j)$. Lastly we sum (3.6) over $k = 1, \dots, K$, and then choose h by minimizing this sum of K martingale residuals. Since the order of such optimal bandwidth is expected to be $n^{-1/5}$ (Fan & Gijbels, 1995), the bandwidth we use for estimation is the minimizer of (3.6) multiplied by a factor n^{-d_0} with $0 < d_0 < 3/10$ such that it is of order n^{-v} with 1/5 < v < 1/2. This assures that the resulting functional estimator $\hat{\Lambda}_v(t_0)$ with the data-dependent smooth parameter has the above desirable large sample properties.

4. EXAMPLE

We illustrate the new proposal with the data combined from three treatment groups of the VALIANT study described in the introduction. Here, the vector U of potentially important baseline risk factors/markers for overall mortality consists of age, Killip class (KC), eGFR, history of

MI (MI), history of congestive heart failure (HF), percutaneous coronary intervention after index MI (CorInt), atrial fibrillation after index MI (AF), history of diabetes (Dia), history of chronic obstructive pulmonary disease (COPD), new left bundle-branch block (BBB), and history of angina (Ang). There are n = 14088 patients (out of a total of 14703) in our analysis, who had complete information about these eleven covariates. The Kaplan-Meier estimate for overall survival for patients in the analysis is given in Figure 1.

First, suppose that we are interested in predicting the 6-month mortality rates of future patients. To this end, we let $t_0 = 6$ (month) and fitted the survival observations truncated slightly after 6 months with a Cox model (2.1) and x = u. The resulting vector of the regression coefficient estimates $\hat{\beta}$ is given in Table 1. These estimates, coupled with the intercept estimate for $\hat{\gamma}$ of -6.02, create a risk score index $\hat{\mathcal{P}}(u)$ for 6 month mortality of future patients.

Now, for a given $\hat{\mathcal{P}}(u) = v$, to obtain $\hat{\tau}(v, 6)$, we let $K(\cdot)$ be the Epanechnikov kernel, and $\psi(v) = \log\{-\log(1-v)\}$ which leads to $\psi\{\hat{\mathcal{P}}(U)\} = \hat{\gamma}' \vec{X}$. The smoothing parameter h = 0.18 was obtained by multiplying a factor of 1/2, which is of order $n^{-0.07}$, to the minimizer of the integrated mean squared martingale residuals defined in (3.6) with 10-fold cross validation. Furthermore, we chose the 2nd and 98th percentiles of the empirical distribution of $\{\hat{\gamma}' \vec{X}_i, i = 1, \dots, n\}$ as the boundary points q_l and q_r for interval \mathcal{J} . To approximate the distribution of (3.1) and W in (3.5), we used the perturbation-resampling method (3.2) with 500 independent realized samples of $\{V_i, i = 1, \dots, n\}$ from the standard normal distribution.

In Figure 2(a), we present a smoothed density estimate for $\hat{\mathcal{P}}(U)$, which provides useful information regarding the relative size of the subgroup of individuals with $\hat{\mathcal{P}}(U) = v$, where $v \in \mathcal{J} =$ [0.02, 0.24]. In Figure 2(b), we present the point estimates $\hat{\tau}(v, 6)$ (solid curve) and 95% pointwise (dark shaded) and simultaneous (gray shaded) interval estimates for $\tau(v, 6)$. The dashed line is 45 degree reference straight line.

Note that the estimated risk scores for most VALIANT patients are less than 0.15. For groups

of subjects with low risk scores, their interval estimates tend to be tight. For example, for subjects with $\hat{\mathcal{P}}(u) = 0.05$, the true 6-month mortality rate is likely to be between 0.03 and 0.05, based on the 95% pointwise confidence interval. The corresponding simultaneous confidence interval is [0.02, 0.06]. For patients whose risk scores are greater than 0.15, the interval estimates are relatively wide as expected. For example, the true mortality rate for patients with $\hat{\mathcal{P}}(u) = 0.20$ is likely to be from 0.19 to 0.25, the boundaries of the 95% pointwise confidence interval. The simultaneous interval is [0.18, 0.27].

Next, suppose that we are interested in predicting long term survival for future patients similar to the VALIANT study population. Here, we let $t_0 = 24$ months and fitted the survival observations slightly truncated at 24 months with a Cox model and x = u. The resulting regression coefficient estimates are also given in Table 1. For this case, the bandwidth h = 0.15 and the range of the estimated risk score v is from 0.04 to 0.48. The smoothed density function estimation of the parametric score is given in Figure 1(c). Note that relatively few patients have scores beyond 0.3. Our consistent point estimates (solid curve) for the true mortality rate and the 95% pointwise (dark region) and simultaneous (gray region) intervals are given in Figure 1 (d). For example, for subjects with a risk score of 0.1, our estimate for their 24 month mortality rate is 0.09 with 95% pointwise and simultaneous confidence intervals being [0.08, 0.11] and [0.07, 0.12], respectively. For patients whose risk scores are high, as expected, their interval estimates can be quite wide. For example, for subjects with risk score of 0.3, the true mortality rate is likely to be between 0.26 and 0.37 based on the 95% simultaneous confidence interval.

5. REMARKS

Due to the complexity of the disease process and the heterogeneity among subjects, it is unlikely that a fitted parametric model would provide a good approximation to the true risk, $P(\tilde{T} < t_0 | U)$, for every subject in the population. The standard method to evaluate the adequacy of an assumed

model is based on a utility or loss function of the observed and predicted responses averaged over the entire study population. When the observed mortality rate is not small in a relatively large sample, it is likely that a standard lack of fit test would reject the fitted model, which can be useful in practice. In the present case, the univariate risk index system constructed from an empirical, semiparametric model can be quite useful for reducing the dimension of the baseline covariate vector and grouping the subjects with "similar" risks. For a given risk scoring scheme, the new proposal presented in this article can be utilized to make valid inferences about the true mortality rate for subjects with the same model-based risk score. On the other hand, to evaluate whether the scoring system effectively partitions the population, one needs to estimate the *distribution* of the risks for these subjects. The index system works well if the distributions are tight around their centers over a range of risk scores of interest. Further research on this challenging problem is warranted.

APPENDIX

In the appendix, we use the standard notations for the empirical process: \mathbb{P}_n and \mathbb{P} represent the expectation with respect to the empirical distribution generated $\{(T_i, \Delta_i, X_i), i = 1, \dots, n\}$ and the distribution generated by (T, Δ, X) , respectively. Similarly, $\mathbb{G}_n = n^{\frac{1}{2}}(\mathbb{P}_n - \mathbb{P})$. Throughout, we assume that $h = O(n^{-\nu})$ with $\frac{1}{5} < \nu < \frac{1}{2}$, K(x) is a symmetric smooth kernel function, has a bounded support [-1, 1] and $\int \dot{K}(x)^2 dx < \infty$, where $\dot{K}(x) = dK(x)/dx$. For convenience, we use the following notations: $K_j(x) = K(x)x^j$, $K_{j,h}(x) = K_j(x/h)/h$ and $m_2 = \int_{-1}^1 K^2(x)dx$. In addition, we assume that the covariate vector X is bounded.

A CONVERGENCE OF $\hat{\mathcal{P}}(U)$

First, we show that the in probability convergence of $\hat{\mathcal{P}}(U) = g(\hat{\gamma}'\vec{X}) \to \bar{\mathcal{P}}(U) = g(\gamma_0'\vec{X})$ for any given U, where $\gamma_0 = (\log\{\Lambda_0(t)\}, \beta_0')', \Lambda_0(t) = \int_0^t \mathbb{P}\{Y(s)\lambda_X(s)\}/\mathbb{P}\{Y(s)e^{\beta_0'X}\}ds], \beta_0 = \operatorname{argmax}_\beta L(\beta),$

$$L(\beta) = \int_0^{t_0} \left[\beta' X - \log \mathbb{P}\{Y(t)e^{\beta' X}\} \right] \mathbb{P}\{Y(t)\lambda_X(t)\}dt,$$
Collection of Boston 12

and $\lambda_x(t)$ is the hazard function of \tilde{T} conditional on X = x. It suffices to show that $\hat{\beta} \to \beta_0$ and $\hat{\Lambda}_0(t) \to \Lambda_0(t)$, in probability. Regardless of the adequacy of the the proportional hazards model, $L(\beta)$ is a convex function and $\hat{L}(\beta) \to L(\beta)$ in probability, uniformly in β . Under the condition that there is no vector ζ such that $\operatorname{pr}(T_1 > T_2 \mid \zeta' X_1 > \zeta' X_2, T_2 \leq t_0) = 1$, $L(\beta)$ has a unique maximizer β_0 . It follows from Theorem 2.1 of Newey & McFadden (1994) that $\hat{\beta} \to \beta_0$ in probability. Furthermore, $|\hat{\beta} - \beta_0| = O_p(n^{-\frac{1}{2}})$. Furthermore, it follows from the same arguments in Lin et al. (1993), $|\hat{\Lambda}_0(t_0) - \Lambda_0(t_0)| = O_p(n^{-\frac{1}{2}})$. This implies the consistency of $\hat{\mathcal{P}}(U)$ for $\bar{\mathcal{P}}(U)$.

B Asymptotic Properties of $\hat{\tau}(v; t_0)$

We next derive the asymptotic properties of $\hat{\tau}(v;t_0)$, for $v \in [\psi^{-1}(\mathfrak{q}_l+h), \psi^{-1}(\mathfrak{q}_r-h)]$. Without the loss of generality, we choose $\psi(x) = \log\{-\log(1-x)\}$. Thus, $\psi\{\bar{\mathcal{P}}(U)\} = \gamma'_0 \vec{X}, \psi\{\hat{\mathcal{P}}(U_i)\} =$ $\hat{\gamma}' \vec{X}_i$, and $\psi\{\hat{\mathcal{P}}(U_i)\} - \psi\{\bar{\mathcal{P}}(U_i)\} = O_p(n^{-\frac{1}{2}})$. Let $\tilde{v} = \psi(v), \lambda_v(t)$ be the hazard function of \tilde{T} conditional on $\gamma'_0 \vec{X} = \tilde{v}, \Lambda_v(t) = \int_0^t \lambda_v(s) ds, \xi(\cdot)$ be the density function of $\gamma'_0 \vec{X}$, and $M_i(t) =$ $N_i(t) - \int_0^t Y_i(s) d\Lambda_{\gamma'_0 \vec{X}_i}(s)$. In view of $\hat{\tau}(v;t_0) = 1 - \exp\{-\hat{\Lambda}_v(t)\}$ and the delta method, in the following we will establish the asymptotical properties of $\hat{\Lambda}_v(t)$. Furthermore, we will repeatedly use the fact that

$$\sup_{e_0 \in [a,b]} \left| \mathbb{P}_n \left[K_h(E-e_0)F - \mathbb{P}\{K_h(E-e_0)F\} \right] \right| = O_p \left[\left\{ (nh)^{-1} \log(n) \right\}^{\frac{1}{2}} \right], \tag{B.1}$$

where $(E_i, F_i), i = 1, \dots, n$ are i.i.d realizations of the random vector (E, F), satisfying that $\sup_{e_0} \mathbb{P}(|F|^s \mid E = e_0) < \infty, \forall s > 0.$

To derive the asymptotic properties of $\hat{\Lambda}_v(t_0)$, we first show that asymptotically there is no extra variation due to the uncertainty in $\hat{\gamma}$. To this end, we write $\hat{\Lambda}_v(t_0) = \hat{\Lambda}_v(t_0, \hat{\gamma})$ and aim to show that for any $\hat{\gamma} = \gamma_0 + O_p(n^{-\frac{1}{2}})$,

$$\sup_{v} |\hat{\Lambda}_{v}(t_{0}, \hat{\gamma}) - \hat{\Lambda}_{v}(t_{0}, \gamma_{0})| = o_{p}\{(nh)^{-\frac{1}{2}}\},$$
(B·2)

where

$$\hat{\Lambda}_{v}(t_{0},\gamma) = \int_{0}^{t_{0}} \frac{\hat{\pi}_{v}^{(2)}(t,\gamma)d\hat{N}_{v}^{(0)}(t,\gamma) - \hat{\pi}_{v}^{(1)}(t,\gamma)d\hat{N}_{v}^{(1)}(t,\gamma)}{\hat{\pi}_{v}^{(2)}(t,\gamma)\hat{\pi}_{v}^{(0)}(t,\gamma) - \hat{\pi}_{v}^{(1)}(t,\gamma)^{2}}$$

$$\hat{N}_{v}^{(k)}(t,\gamma) = n^{-1}\sum_{i=1}^{n} K_{h}\{D_{vi}(\gamma)\}\{h^{-1}D_{vi}(\gamma)\}^{k}N_{i}(t),$$

$$\hat{\pi}_{v}^{(k)}(t,\gamma) = n^{-1}\sum_{i=1}^{n} K_{h}\{D_{vi}(\gamma)\}\{h^{-1}D_{vi}(\gamma)\}^{k}Y_{i}(t),$$
(B·3)

and $D_{vi}(\gamma) = \psi\{g(\gamma' \vec{X_i})\} - \psi(v)$. Here and in the sequel, the sup is taken over $[\psi^{-1}(\mathfrak{q}_l + h), \psi^{-1}(\mathfrak{q}_r - h)]$ for v and over $[0, t_0]$ for t. To this end, we write

$$\hat{\Lambda}_{v}(t_{0},\gamma) = \int_{0}^{t_{0}} \left\{ \hat{R}_{v}^{(2)}(t,\gamma) d\hat{N}_{v}^{(0)}(t,\gamma) - \hat{R}_{v}^{(1)}(t,\gamma) d\hat{N}_{v}^{(1)}(t,\gamma) \right\}$$

and seek to show that

$$\sup_{t,v} \left| \hat{\pi}_v^{(k)}(t,\hat{\gamma}) - \hat{\pi}_v^{(k)}(t,\gamma_t) \right| = o_p\{(nh)^{-\frac{1}{2}}\}.$$
 (B·4)

where

$$\hat{R}_{v}^{(k)}(t,\gamma) = \frac{\hat{\pi}_{v}^{(k)}(t,\gamma)}{\hat{\pi}_{v}^{(2)}(t,\gamma)\hat{\pi}_{v}^{(0)}(t,\gamma) - \hat{\pi}_{v}^{(1)}(t,\gamma)^{2}}, \quad k = 1, 2.$$

Since $\hat{\pi}_v^{(k)}(t,\gamma) = \int K_{j,h}(x-\tilde{v}) d\mathbb{P}_n \{ I(\hat{\gamma}' \vec{X} \le x) Y(t) \},\$

$$\begin{aligned} &|\hat{\pi}_{v}^{(k)}(t,\hat{\gamma}) - \hat{\pi}_{v}^{(k)}(t,\gamma_{0})| = \left| \int K_{j,h}(x-\tilde{v})d\mathbb{P}_{n}\{I(\hat{\gamma}'\vec{X} \leq x)Y(t) - I(\gamma_{0}'\vec{X} \leq x)Y(t)\}| \right| \\ &\leq \left| n^{-\frac{1}{2}} \int K_{j,h}(x-\tilde{v})d\mathbb{G}_{n}\{I(\hat{\gamma}'\vec{X} \leq x)Y(t) - I(\gamma_{0}'\vec{X} \leq x)Y(t)\}| \right| \\ &+ \left| \int K_{j,h}(x-\tilde{v})\mathbb{P}\{I(\hat{\gamma}'\vec{X} \leq x)Y(t) - I(\gamma_{0}'\vec{X} \leq x)Y(t)\}| \right| \\ &\leq h^{-1}n^{-\frac{1}{2}} ||\mathbb{G}_{n}||_{\mathcal{H}_{\delta}} + O_{p}(n^{-\frac{1}{2}}) \end{aligned}$$

where $\mathcal{H}_{\delta} = \{ [I(\gamma'\vec{X} \leq x) - I(\gamma'_{0}\vec{X} \leq x)]Y(t) : x, s, |\gamma - \gamma_{0}| \leq \delta \}$ is a class of functions indexed by x, s and δ . Furthermore, \mathcal{H}_{δ} is uniformly bounded by an envelop function in the order of $\delta^{\frac{1}{2}}$ with respect to L_{2} norm. By the maximum inequality (Theorem 2.14.2, Van der Vaart and Wellner, 1996) and the fact that $|\hat{\gamma} - \gamma_{0}| = O_{p}(n^{-\frac{1}{2}})$, we have

$$\begin{split} & \text{Collection of Biostatistics} \\ & \text{Research Archive} \\ \end{split} \\ \begin{array}{l} & h^{-1}n^{-\frac{1}{2}} \|\mathbb{G}_n\|_{\mathcal{H}_{\delta}} \lesssim O_p\{h^{-1}n^{-\frac{1}{2}}n^{-\frac{1}{4}}\log(n)\}. \end{split}$$

It follows that $|\hat{\pi}_v^{(k)}(t,\hat{\gamma}) - \hat{\pi}_v^{(k)}(t,\gamma_0)| = (nh)^{-\frac{1}{2}}(nh^2)^{-\frac{1}{4}}\log(n) + O_p(n^{-\frac{1}{2}}) = o_p\{(nh)^{-\frac{1}{2}}\}.$ This, together with $(B \cdot 1)$, implies that

$$\sup_{t,v} |\hat{R}_v^{(k)}(t,\gamma_0) - \hat{R}_v^{(k)}(t,\hat{\gamma})| = o_p\{(nh)^{-\frac{1}{2}}\}, \quad \mathbf{k} = 0, 1,$$

and
$$\sup_{t,v}\{|\hat{R}_v^{(2)}(t,\gamma_0) - r_v^{(2)}(t)| + |\hat{R}_v^{(1)}(t,\gamma_0)|\} = O_p\{(nh)^{-\frac{1}{2}}\log(n)^{\frac{1}{2}} + h^2\},$$

where $r_v^{(2)}(t) = 1/\{\xi(\tilde{v})\mathcal{S}_v(t)\}$ and $\mathcal{S}_v(t) = \operatorname{pr}(T \ge t|\gamma_0'\vec{X} = \tilde{v}).$

Next, we note that

$$\left| \int_{0}^{t_{0}} \hat{R}_{v}^{(2)}(t,\hat{\gamma}) d\hat{N}_{v}^{(0)}(t,\hat{\gamma}) - \int_{0}^{t_{0}} \hat{R}_{v}^{(2)}(t,\gamma_{0}) d\hat{N}_{v}^{(0)}(t,\gamma_{0}) \right| \leq \hat{\epsilon}_{1} + \hat{\epsilon}_{2} + \hat{\epsilon}_{3}$$

where

$$\begin{aligned} \widehat{\epsilon}_{1} &= \mathbb{P}_{n} \left[\int_{0}^{t_{0}} \left| \hat{R}_{v}^{(2)}(t,\hat{\gamma}) - \hat{R}_{v}^{(2)}(t,\gamma_{0}) \right| K_{h}(\hat{\gamma}'\vec{X} - \tilde{v}) dN(t) \right] + \\ \widehat{\epsilon}_{2} &= \mathbb{P}_{n} \left| \left\{ K_{h}(\hat{\gamma}'\vec{X} - \tilde{v}) - K_{h}(\gamma_{0}'\vec{X} - \tilde{v}) \right\} \int_{0}^{t_{0}} \left| \hat{R}_{v}^{(2)}(t,\gamma_{0}) - r_{v}^{(2)}(t) \right| dN(t) \right| \\ \widehat{\epsilon}_{3} &= \left| \mathbb{P}_{n} \left[\left\{ K_{h}(\hat{\gamma}'\vec{X} - \tilde{v}) - K_{h}(\gamma_{0}'\vec{X} - \tilde{v}) \right\} \int_{0}^{t_{0}} r_{v}^{(2)}(t) dN(t) \right\} \right] \right| \end{aligned}$$

We will bound these three terms in the following. For the first term,

$$\widehat{\epsilon}_1 \le o_p\{(nh)^{-\frac{1}{2}}\}\mathbb{P}_n\left\{K_h(\widehat{\gamma}'\vec{X} - \widetilde{v})N(t_0)\right\} = o_p\{(nh)^{-\frac{1}{2}}\}.$$

For the second term,

$$\begin{aligned} \widehat{\epsilon}_{2} &\leq h^{-1} \sup_{t} \left| \hat{R}_{v}^{(2)}(t,\gamma_{0}) - r_{v}^{(2)}(t) \right| \sup_{x} \left[\mathbb{P}_{n} \left\{ \left| I(\hat{\gamma}'\vec{X} \leq x) - I(\gamma_{0}'\vec{X} \leq x) \right| N(t_{0}) \right\} \right] \\ &\leq O_{p} \{ h^{-1}(nh)^{-\frac{1}{2}} \log(n)^{\frac{1}{2}} + h \} O_{p}(n^{-\frac{1}{2}}) = o_{p} \{ (nh)^{-\frac{1}{2}} \}. \end{aligned}$$

For the last term,

$$\begin{aligned} \widehat{\epsilon}_3 &\leq n^{-\frac{1}{2}} h^{-1} \sup_x \left| \mathbb{G}_n \left[\left\{ I(\widehat{\gamma}' \vec{X} \leq x) - I(\gamma_0' \vec{X} \leq x) \right\} \int_0^{t_0} r_v^{(2)}(t) dN(t) \right] \right| + O_p(n^{-\frac{1}{2}}) \\ &= O_p\{ n^{-\frac{1}{2}} h^{-1} n^{-\frac{1}{4}} \log(n) \} = o_p\{ (nh)^{-\frac{1}{2}} \}. \end{aligned}$$

Therefore,

$$\sup_{v} \left| \int_{0}^{t_{0}} \hat{R}_{v}^{(2)}(t,\hat{\gamma}) d\hat{N}_{v}^{(0)}(t,\hat{\gamma}) - \int_{0}^{t_{0}} \hat{R}_{v}^{(2)}(t,\gamma_{0}) d\hat{N}_{v}^{(0)}(t,\gamma_{0}) \right| = o_{p}\{(nh)^{-\frac{1}{2}}\}.$$

Similarly,

$$\int_0^{t_0} \hat{R}_v^{(1)}(t,\hat{\gamma}) d\hat{N}_v^{(1)}(t,\hat{\gamma}) - \int_0^{t_0} \hat{R}_v^{(1)}(t,\gamma_t) d\hat{N}_v^{(1)}(t,\gamma) = o_p\{(nh)^{-\frac{1}{2}}\},$$

and thus (B·2) holds. On the other hand, the arguments given in Li & Doss (1995) can be used to show that $\sup_{v} |\hat{\Lambda}_{v}(t_{0}, \gamma_{0}) - \Lambda_{v}(t_{0})| \to 0$, in probability. It follows that $\sup_{v} |\hat{\Lambda}_{v}(t_{0}) - \Lambda_{v}(t_{0})| \leq \sup_{v} |\hat{\Lambda}_{v}(t_{0}, \hat{\gamma}) - \hat{\Lambda}_{v}(t_{0}, \gamma_{0})| + \sup_{v} |\hat{\Lambda}_{v}(t_{0}, \gamma_{0}) - \Lambda_{v}(t_{0})| \to 0$, in probability. This concludes the uniform consistency of $\hat{\Lambda}_{v}(t_{0})$.

We next derive the asymptotic distribution of $\widehat{\mathcal{W}}_v(t_0) = (nh)^{\frac{1}{2}} \{ \widehat{\Lambda}_v(t_0) - \Lambda_v(t_0) \}$. From (B·2), we have $\widehat{\mathcal{W}}_v(t_0) = \widetilde{\mathcal{W}}_v(t_0) + o_p(1)$, where $\widetilde{\mathcal{W}}_v(t_0) = (nh)^{\frac{1}{2}} \{ \widehat{\Lambda}_v(t_0, \gamma_0) - \Lambda_v(t_0) \}$. Noting that $n^{\frac{1}{2}}h^{\frac{5}{2}} = o_p(1)$ and the decomposition that

$$\begin{split} \widetilde{\mathcal{W}}_{v}(t_{0}) &= (nh)^{\frac{1}{2}} \mathbb{P}_{n} \left[\int_{0}^{t_{0}} \left\{ \hat{R}_{v}^{(2)}(t,\gamma_{0}) K_{h}(\gamma_{0}'\vec{X}-\tilde{v}) - \hat{R}_{v}^{(1)}(t,\gamma_{0}) K_{1,h}(\gamma_{0}'\vec{X}-\tilde{v}) \right\} dM(t) \right] \\ &+ (nh)^{\frac{1}{2}} \left\{ \frac{h^{2} + o_{p}(h^{2})}{2} \int_{0}^{t_{0}} \frac{\hat{\pi}_{v}^{(2)}(t,\gamma_{0})^{2} - \hat{\pi}_{v}^{(1)}(t,\gamma_{0}) \hat{\pi}_{v}^{(3)}(t,\gamma_{0})}{\hat{\pi}_{v}^{(2)}(t,\gamma_{0}) \hat{\pi}_{v}^{(0)}(t,\gamma_{0}) - \hat{\pi}_{v}^{(1)}(t,\gamma_{0})^{2}} \frac{\partial^{2} \lambda_{v}(t)}{\partial \tilde{v}^{2}} ds \right\}, \end{split}$$

we have

$$\widetilde{\mathcal{W}}_{v}(t_{0}) = (nh)^{\frac{1}{2}} \mathbb{P}_{n} \left[\int_{0}^{t_{0}} \left\{ \hat{R}_{v}^{(2)}(t,\gamma_{0}) K_{h}(\gamma_{0}'\vec{X}-\tilde{v}) - \hat{R}_{v}^{(1)}(t,\gamma_{0}) K_{1,h}(\gamma_{0}'\vec{X}-\tilde{v}) \right\} dM(t) \right] + o_{p}(1).$$

By integration by part and maximum inequality for empirical process, we have

$$(nh)^{\frac{1}{2}}\mathbb{P}_n\left\{K_{1,h}(\gamma_0'\vec{X}-\tilde{v})\int_0^{t_0}\hat{R}_v^{(1)}(t,\gamma_0)Y(t)dM(t)\right\} = o_p(1)$$

and

$$(nh)^{\frac{1}{2}}\mathbb{P}_n\left[K_h(\gamma_0'\vec{X}-\tilde{v})\int_0^{t_0}\left\{\hat{R}_v^{(2)}(t,\gamma_0)-r_v^{(2)}(t)\right\}dM(t)\right]=o_p(1),$$

uniformly in \tilde{v} . Therefore,

$$\widehat{\mathcal{W}}_{v}(t) = (nh)^{\frac{1}{2}} \mathbb{P}_{n} \left\{ K_{h}(\gamma_{0}' \vec{X} - \tilde{v}) \int_{0}^{t_{0}} r_{v}^{(2)}(t) dM(t) \right\} + o_{p}(1).$$
(B·5)

For any fixed v (or \tilde{v}), by martingale central limit theorem,

$$(nh)^{\frac{1}{2}} \mathbb{P}_n \left\{ K_h(\gamma'_0 \vec{X} - \tilde{v}) \int_0^{t_0} r_v^{(2)}(t) dM(t) \right\} \to N\{0, \sigma_v^2(t_0)\},$$

in distribution, as $n \to \infty$, where

$$\sigma_v^2(t_0) = m_2 \int_0^{t_0} r_v^{(2)}(t) d\Lambda_v(t)$$

It follows that for any fixed v, $\widehat{\mathcal{W}}_v(t)$ converges in distribution to $N\{0, \sigma_v^2(t_0)\}$.

To demonstrate the validity of the resampling variance estimator, we define

$$\hat{\gamma}^* = (\log\{\hat{\Lambda}_0^*(t_0)\}, \hat{\beta}^{*\prime})',$$
 (B·6)

where

$$\hat{\beta}^{*} = \hat{\beta} + \hat{\mathbb{A}}^{-1} \sum_{i=1}^{n} \int \left[V_{i} \left\{ X_{i} - \frac{\hat{\Pi}^{(1)}(t,\hat{\beta})}{\hat{\Pi}^{(0)}(t,\hat{\beta})} \right\} - \frac{n^{-1} \sum_{j=1}^{n} V_{j} Y_{j}(t) e^{\hat{\beta}' X_{j}} \left\{ \hat{\Pi}^{(0)}(t,\hat{\beta}) X_{j} - \hat{\Pi}^{(1)}(t,\hat{\beta}) \right\}}{\hat{\Pi}^{(0)}(t,\hat{\beta})^{2}} \right] dN_{i}(t)$$

$$\log\{\hat{\Lambda}^{*}_{0}(t_{0})\} = \log\{\hat{\Lambda}_{0}(t_{0})\} + \frac{n^{-1}}{\hat{\Lambda}_{0}(t)} \sum_{i=1}^{n} \int_{0}^{t} \left[\frac{V_{i}}{\hat{\Pi}^{(0)}(s,\hat{\beta})} - \frac{n^{-1} \sum_{j=1}^{n} V_{j} Y_{j}(s) e^{\hat{\beta}' X_{j}} + \hat{\Pi}^{(1)}(t,\hat{\beta})'(\hat{\beta}^{*} - \hat{\beta})}{\hat{\Pi}^{(0)}(s,\hat{\beta})^{2}} \right] dN_{i}(s)$$

 $\hat{\Pi}^{(k)}(t,\beta) = n^{-1} \sum_{j=1}^{n} Y_j(t) e^{\beta' X_j} X_j^{\otimes k}$ and for any vector $x, x^{\otimes 0} = 1, x^{\otimes 1} = x$ and $x^{\otimes 2} = xx'$. It is straightforward to show that the distribution of $n^{\frac{1}{2}}(\hat{\gamma}^* - \hat{\gamma})$ conditional on the data and the unconditional distribution of $n^{\frac{1}{2}}(\hat{\gamma} - \gamma_0)$ converge to the same limiting normal distribution. Now since $\hat{\gamma}^* = \gamma_0 + O_p(n^{-\frac{1}{2}})$, by (B·2),

$$(3.2) = n^{-\frac{1}{2}} h^{\frac{1}{2}} \sum_{i=1}^{n} K_h(\hat{D}_{vi}) \hat{\mathcal{V}}_i(v; t_0) V_i + o_p(1),$$

Thus the variance estimator for $(nh)^{\frac{1}{2}} \{ \hat{\Lambda}_v(t_0) - \Lambda_v(t_0) \}$ is

$$\begin{split} \hat{\sigma}_{v}^{2}(t_{0}) = &\operatorname{var}\left[n^{-\frac{1}{2}}h^{\frac{1}{2}}\sum_{i=1}^{n}K_{h}(\hat{D}_{vi})\hat{\mathcal{V}}_{i}(v;t_{0})V_{i} \middle| (T_{i},\Delta_{i},U_{i}), i = 1,\cdots,n\right] + o_{p}(1) \\ &= \mathbb{P}_{n}\left[hK_{h}^{2}\left(\hat{\gamma}'\vec{X} - \tilde{v}\right)\left\{\int_{0}^{t_{0}}\frac{dN(t) - Y(t)d\hat{\Lambda}_{v}(t)}{\hat{\pi}_{v}^{(0)}(t,\hat{\gamma})}\right\}^{2}\right] + o_{p}(1) \\ &= \mathbb{P}_{n}\left[hK_{h}^{2}\left(\gamma_{0}'\vec{X} - \tilde{v}\right)\left\{\int_{0}^{t_{0}}r_{v}^{(2)}(t)dM(t)\right\}^{2}\right] + o_{p}(1) \\ &= \mathbb{P}\left\{hK_{h}^{2}\left(\gamma_{0}'\vec{X} - \tilde{v}\right)\int_{0}^{t_{0}}r_{v}^{(2)}(t)^{2}Y(t)d\Lambda_{v}(t)\right\} + o_{p}(1) = \sigma_{v}^{2}(t_{0}) + o_{p}(1) \end{split}$$

which converges to $\sigma_v^2(t_0)$ in probability.

C – Justification for the Validity of the Confidence Band for $\hat{ au}(v;t_0)$

We first justify that after proper standardization, the supermum type statistics

$$W = \sup_{v} \left| \frac{(nh)^{\frac{1}{2}} \hat{\Lambda}_v(t_0) - \Lambda_v(t_0)}{\hat{\sigma}_v(t_0)} \right|$$

converges weakly. It follows from (B.5) and the consistency of $\hat{\sigma}_v(t_0)$ for $\sigma_v(t_0)$ that

$$W = \sup_{v} \left| (nh)^{\frac{1}{2}} \mathbb{P}_{n} \left\{ K_{h}(\gamma_{0}'\vec{X} - \tilde{v}) \int_{0}^{t_{0}} \frac{r_{v}^{(2)}(t)}{\sigma_{v}(t_{0})} dM_{i}(t) \right\} \right| + o_{p}(1).$$

This, together with the continuity of $r_v^{(2)}(t)/\sigma_v(t_0)$, implies that

$$W = \sup_{v} \left| (nh)^{\frac{1}{2}} \mathbb{P}_n \left\{ K_h(\gamma'_0 \vec{X} - \tilde{v}) \eta \right\} \right| + o_p(1).$$

where

$$\eta = \int_0^{t_0} \frac{r_{\psi^{-1}(\gamma'_0 \vec{X})}^{(2)}(t) dM(t)}{\sigma_{\psi^{-1}(\gamma'_0 \vec{X})}(t_0)}$$

If follows from the similar argument in Bickel & Rosenblatt (1973) that

$$\operatorname{pr}\left\{a_n(W - d_n) < x\right\} \to e^{-2e^{-x}}$$

where

$$a_n = \left\{ 2 \log \left(\frac{\psi(\mathbf{q}_r) - \psi(\mathbf{q}_l)}{h} \right) \right\}^{\frac{1}{2}} \quad \text{and} \quad d_n = a_n + a_n^{-1} \log \left\{ \frac{1}{4m_2 \pi} \int \dot{K}(t)^2 dt \right\}.$$

Now, to justify the validity of the resampling procedure for constructing the confidence band, we consider the statistics

$$W^* = \sup_{v} \left| \frac{(nh)^{\frac{1}{2}}}{\hat{\sigma}_v(t_0)} \sum_{i=1}^n V_i \int_0^{t_0} \frac{K_h(\hat{\gamma}'\vec{X}_i - \tilde{v}) \{ dN_i(t) - Y_i(t) d\hat{\Lambda}_v(t) \}}{\hat{\pi}_v^{(0)}(t,\hat{\gamma})} + (nh)^{\frac{1}{2}} \{ \hat{\Lambda}_v(t_0, \hat{\gamma}^*) - \hat{\Lambda}_v(t_0, \hat{\gamma}) \} \right|,$$

Again, since $|\hat{\gamma}^* - \hat{\gamma}| = O_p(n^{-\frac{1}{2}})$, from (B·2), we have

$$W^* = \sup_{v} \left| \frac{(nh)^{\frac{1}{2}}}{\hat{\sigma}_v(t_0)} \sum_{i=1}^n V_i \int_0^{t_0} \frac{K_h(\hat{\gamma}'\vec{X}_i - \tilde{v}) \left\{ dN_i(t) - Y_i(t) d\hat{\Lambda}_v(t) \right\}}{\hat{\pi}_v^{(0)}(t, \hat{\gamma})} \right| + o_p(1)$$

It follows from the same argument as given in Tian et al. (2005) that,

$$\sup_{x} \left| \Pr \left\{ a_n (W^* - d_n) < x \mid (T_i, \Delta_i, U_i), i = 1, \cdots, n \right\} - e^{-2e^{-x}} \right|$$

 $\rightarrow 0$, in probability as $n \rightarrow \infty$. Therefore, the conditional distribution of $a_n(W^* - d_n)$ can be used to approximate the distribution of $a_n(W - d_n)$ for large n.



References

- ANAVEKAR, N., MCMURRAY, J., VELAZQUEZ, E., SOLOMON, S., KOBER, L., ROULEAU, J., WHITE, H., NORDLANDER, R., MAGGIONI, A., DICKSTEIN, K. ET AL. (2004). Relation between Renal Dysfunction and Cardiovascular Outcomes after Myocardial Infarction.
- ANDERSON, K., WILSON, P., ODELL, P. & KANNEL, W. (1991). An updated coronary risk profile. A statement for health professionals. *Circulation* 83, 356–362.
- BICKEL, P. J. & ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates (Corr: V3 p1370). *The Annals of Statistics* 1, 1071–1095.
- BRINDLE, P., EMBERSON, J., LAMPE, F., WALKER, M., WHINCUP, P., FAHEY, T. & EBRAHIM, S. (2003). Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study.
- CAI, T., TIAN, L. & WEI, L. (2005). Semiparametric Box-Cox power transformation models for censored survival observations. *Biometrika* 92, 619–632.
- COOK, N. (2007). Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. *Circulation* **115**, 928.
- DABROWSKA, D. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. Ann. Statist 17, 1157–1167.
- D'AGOSTINO, SR, R., GRUNDY, S., SULLIVAN, L. & WILSON, P. (2001). Validation of the Framingham Coronary Heart Disease Prediction Scores Results of a Multiple Ethnic Groups Investigation.
- DU, Y. & AKRITAS, M. (2002). Iid representations of the conditional Kaplan-Meier process for arbitrary distributions. *Math. Methods Statist* 11, 152–182.
- FAN, J. & GIJBELS, I. (1995). Data-driven bandwidth selection in local polynomial regression: variable bandwidth selection and spatial adaptation. *Journal of the Royal Statistical Association*, *Series B* 57, 371–394.

FAN, J. & GIJBELS, I. (1996). Local polynomial modelling and its applications, Vol. 66 of Monographs on Statistics and Applied Probability. *London: Chapman Hall*, .

HÄRDLE, W. (1990). Applied Nonparametric Regression. Cambridge University Press.

- JIN, Z., YING, Z. & WEI, L. (2001). A simple resampling method by perturbing the minimand. Biometrika 88, 381–390.
- KALBFLEISCH, J. D. & PRENTICE, R. L. (2002). The Statistical Analysis of Failure Time Data. John Wiley & Sons.
- LI, G. & Doss, H. (1995). An approach to nonparametric regression for life history data using local linear fitting. Ann. Statist 23, 787–823.
- LIN, D., WEI, L. & YING, Z. (1993). Checking the Cox model with cumulative sums of martingalebased residuals. *Biometrika* **80**, 557–572.
- MAMMEN, E. (1992). Bootstrap, wild bootstrap, and asymptotic normality. *Probability Theory* and Related Fields **93**, 439–455.
- NEWEY, W. & MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. In Handbook of Econometrics, Ed. D. McFadden and R. Engler, volume IV, pp. 2113–245. Amsterdam: North Holland.
- 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.
- PEPE, M., JANES, H., LONGTON, G., LEISENRING, W. & NEWCOMB, P. (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology* 159, 882–890.
- PFEFFER, M., MCMURRAY, J., VELAZQUEZ, E., ROULEAU, J., KOBER, L., MAGGIONI, A., Collection of Biostatistics

SOLOMON, S., SWEDBERG, K., VAN DE WERF, F., WHITE, H. ET AL. (2003). Valsartan, Captopril, or Both in Myocardial Infarction Complicated by Heart Failure, Left Ventricular Dysfunction, or Both.

- SOLOMON, S., SKALI, H., ANAVEKAR, N., BOURGOUN, M., BARVIK, S., GHALI, J., WARNICA, J., KHRAKOVSKAYA, M., ARNOLD, J., SCHWARTZ, Y. ET AL. (2005). Changes in Ventricular Size and Function in Patients Treated With Valsartan, Captopril, or Both After Myocardial Infarction.
- 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.
- UNO, H., CAI, T., TIAN, L. & WEI, L. J. (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). Transformation in density estimation (with comments). Journal of the American Statistical Association 86, 343–361.
- WANG, T., GONA, P., LARSON, M., TOFLER, G., LEVY, D., NEWTON-CHEH, C., JACQUES, P., RIFAI, N., SELHUB, J., ROBINS, S. ET AL. (2006). Multiple Biomarkers for the Prediction of First Major Cardiovascular Events and Death. *New England Journal of Medicine* 355, 2631.
- WARE, J. (2006). The Limitations of Risk Factors as Prognostic Tools. New England Journal of Medicine 355, 2615.
- WU, C. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. Annals of Statistics 14, 1261–1295.
- ZHANG, X., ATTIA, J., D'ESTE, C., YU, X. & WU, X. (2005). A risk score predicted coronary heart disease and stroke in a Chinese cohort. *Journal of Clinical Epidemiology* 58, 951–958.



Figure 1: The Kaplan-Meier Curve with the survival time data for the entire VALIANT study.



Table 1: Estimates (Est) of the regression coefficients, the corresponding standard errors (SE) and p-values by fitting the Cox model to the VALIANT data based on survival time information up to t_0 .

		Covariates										
		Age	KC	eGFR	MI	HF	CorInt	AF	Dia	COPD	BBB	Ang
	Est	.028	.289	012	.206	.242	433	.252	.241	.161	.220	.164
$t_0 = 6$	SE	.003	.032	.002	.066	.072	.107	.072	.061	.088	.108	.062
	p-value	.000	.000	.000	.002	.001	.000	.001	.000	.070	.042	.008
	Est	.030	.237	010	.295	.327	463	.296	.268	.243	.231	.235
$t_0 = 24$	SE	.002	.023	.001	.047	.051	.077	.052	.044	.062	.080	.045
	p-value	.000	.000	.000	.000	.000	.000	.000	.000	.000	.004	.000



Figure 2: Prediction of 6-month and 24-month mortality risks based on the eleven risk factors. The top panel shows the estimated density function of the model based risk estimate for (a) $t_0 = 6$ months; and (c) $t_0 = 24$ months. The lower panel shows the point (thick solid curve) estimate of the true risk function $\tau(v, t_0)$ along with their pointwise (dark shaded region) and simultaneous (light shaded region) confidence intervals for (b) $t_0 = 6$ months; and (d) $t_0 = 24$ months. The dashed line is the 45 degree reference line.





(a) Density for 6 month mortality risk

(c) Estimated Density for 24 month mortality risk



