Calculating the Concordance Probability Estimate with a Discrete Number of Risk Groups

Glenn Heller,[1,*] **Qianxing Mo,**[2]

[1] Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, U.S.A.

[2] Department of Medicine and Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX 77030, U.S.A.

**email:* hellerg@mskcc.org

SUMMARY. The concordance probability estimate (CPE) is used to determine the discriminatory power of the proportional hazards model. Previously, the CPE was developed for continuous relative risk scores in the proportional hazards model. In the current paper, the CPE is modified to account for ties in the risk scores. The development of clinical staging systems is an important application of model development with a discrete number of risk scores. The asymptotic distribution of the CPE with ties is derived. Simulations are generated and the CPE is compared to the c-index. An example exploring the strength of a risk classification system for metastatic prostate cancer strength is used to illustrate the methodology.

KEY WORDS: Concordance; Discrimination; Proportional hazards; Survival analysis; Staging

1. Introduction

In clinical research, statistical models to assess patient risk are often utilized to assist in treatment decisions and prognosis. The quality of these decisions is tied to the adequacy and the strength of the model. In this paper, we consider the utilization of the Cox proportional hazards risk model with baseline covariates to determine the patient risk of death (Cox, 1972). It is assumed that the adequacy of the proportional hazards assumptions has been empirically tested and that the model assumptions cannot be rejected. Two common methods used to test the proportional hazards assumption are developed in Lin et al. (1993) and Grambsch and Therneau (1994).

A commonly employed measure of strength for the proportional hazards model is its discriminatory power, a metric used to delineate high risk and low risk patients. For the proportional hazards model with independent right censored data, a consistent estimate for discrimination is the concordance probability estimate (Gönen and Heller, 2005). The proportional hazards model is denoted by

$$h(t|\boldsymbol{x}) = h_0(t)\exp[\boldsymbol{\beta}^T\boldsymbol{x}],$$

where t represents survival time and $\boldsymbol{\beta}^T\boldsymbol{x}$ is the patient-specific relative risk score based on the covariate vector $\boldsymbol{x}$. Under the proportional hazards specification, the concordance probability $\Pr(T_2 > T_1|\boldsymbol{\beta}^T\boldsymbol{x}_1 > \boldsymbol{\beta}^T\boldsymbol{x}_2)$ is equal to

$$\frac{\iint\limits_{\boldsymbol{\beta}^T\boldsymbol{x}_1>\boldsymbol{\beta}^T\boldsymbol{x}_2} \left[1+\exp\left\{\boldsymbol{\beta}^T(\boldsymbol{x}_2-\boldsymbol{x}_1)\right\}\right]^{-1} dF(\boldsymbol{\beta}^T\boldsymbol{x}_1)dF(\boldsymbol{\beta}^T\boldsymbol{x}_2)}{\iint\limits_{\boldsymbol{\beta}^T\boldsymbol{x}_1>\boldsymbol{\beta}^T\boldsymbol{x}_2} dF(\boldsymbol{\beta}^T\boldsymbol{x}_1)dF(\boldsymbol{\beta}^T\boldsymbol{x}_2)}$$

where F is the distribution function of the relative risk score. The concordance probability is estimated by substituting the partial likelihood estimate of $\boldsymbol{\beta}$ (Cox, 1975)

and using the empirical distribution function for F. The result is the concordance probability estimate (CPE)

$$\frac{2}{n(n-1)} \sum\sum_{i<j} \left\{ \frac{I(\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_{ji} < 0)}{1 + \exp(\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_{ji})} + \frac{I(\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_{ij} < 0)}{1 + \exp(\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_{ij})} \right\}$$

where $\boldsymbol{x}_{ij}$ represents the pairwise difference $\boldsymbol{x}_i - \boldsymbol{x}_j$. In addition to its consistency, it was previously shown through simulation that with moderate sample size the CPE is relatively unaffected by independent censoring and is therefore a robust measure of discrimination with survival data.

The CPE was developed for continuous relative risk scores. The indicator functions used for its computation are determined with strict inequalities. There are, however, many clinical scenarios, like the development of clinical staging systems, where there are ties in the risk scores. In oncology, staging systems are widespread and the American Joint Committee on Cancer have developed and revised staging systems for solid tumors, lymphomas and leukemias (Edge et al., 2010). The creation of clinical stages induces a large number of ties in the risk scores. In this paper, the CPE is modified to incorporate ties in the risk scores.

An earlier version of the concordance probability estimate, instituted in the `R` package `CPE`, included ties in the risk scores using the calculation

$$\frac{2}{n(n-1)} \sum\sum_{i<j} \left\{ \frac{I(\hat{\boldsymbol{\beta}}^T x_{ji} \leq 0)}{1 + \exp(\hat{\boldsymbol{\beta}}^T x_{ji})} + \frac{I(\hat{\boldsymbol{\beta}}^T x_{ij} < 0)}{1 + \exp(\hat{\boldsymbol{\beta}}^T x_{ij})} \right\}.$$

This had the effect of adding 1/2 for patient pairs with tied risk scores and when applied to data with a large number of ties, resulted in a considerable attenuation of the CPE. In the next section, we revert to the original definition of the concordance probability $\Pr(T_2 > T_1 | \boldsymbol{\beta}^T \boldsymbol{x}_1 > \boldsymbol{\beta}^T \boldsymbol{x}_2)$, and compute the CPE by explicitly removing

patient pairs with tied risk scores. The updated calculation creates a statistic with a random numerator and denominator. The asymptotic distribution of the CPE is derived through the ratio of U statistics.

2. CPE by removing ties in the risk scores

The CPE excluding ties in the risk scores is

$$K_n(\hat{\boldsymbol{\beta}}) = \frac{\sum\sum_{i<j} \left\{ \frac{I(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_{ji}<0)}{1+\exp(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_{ji})} + \frac{I(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_{ij}<0)}{1+\exp(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_{ij})} \right\}}{\sum\sum_{i<j} \{I(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_{ji} < 0) + I(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}_{ij} < 0)\}},$$

where $\hat{\boldsymbol{\beta}}$ is the partial likelihood estimate of $\boldsymbol{\beta}$ from the proportional hazards model. Under the proportional hazards specification, $K_n(\hat{\boldsymbol{\beta}})$ converges weakly to the concordance probability $K(\boldsymbol{\beta}) = \Pr(T_2 > T_1 | \boldsymbol{\beta}^T\boldsymbol{x}_1 > \boldsymbol{\beta}^T\boldsymbol{x}_2)$. The asymptotic distribution of $K_n(\hat{\boldsymbol{\beta}})$ is given in the theorem below.

THEOREM 1. Under the standard conditions for the proportional hazards model, $n^{1/2}[K_n(\hat{\boldsymbol{\beta}}) - K(\boldsymbol{\beta})]$ is asymptotically normal with mean 0 and variance $V(\boldsymbol{\beta})$. The asymptotic variance and its estimate are derived in the appendix.

3. Simulations

Simulation experiments were conducted to examine the accuracy of the CPE for a four group staging system. The data were generated from a Weibull regression model

$$T_i = \exp[0.5x_{1i} + 0.25x_{2i} + 0.10x_{3i}] \times \epsilon_i$$

where $\boldsymbol{x} = (x_1, x_2, x_3)^T$ are dummy variables. The errors $\{\epsilon_i\}$ were independent

identically distributed Weibull random variables with scale parameter 1 and shape parameters $\{1.85, 4.1, 7.3, 13.5\}$, which were chosen to produce concordance probabilities equal to $\{0.6, 0.7, 0.8, 0.9\}$. The Weibull regression model enables the data to satisfy the proportional hazards assumption. For each group, independent uniform censoring times $(0, \tau_k), (k = 1, 2, 3, 4)$, were generated to determine the proportion censored. The sample size for each simulation was 200 and the proportion of patients in the four groups was $\{0.1, 0.2, 0.3, 0.4\}$. One thousand simulations were run to produce the results in Table 1.

In addition to the CPE, the c-index with inverse probability censoring weights was included in the simulations for comparison (Harrell et al. 1982, Uno et al. 2011). The inverse probability censoring weights are incorporated to produce an asymptotically unbiased statistic when the support of the survival distribution is less than the support of the censoring distribution. The c-index with inverse probability censoring weights is defined as

$$c = \frac{\sum_i \sum_j I(y_i < y_j) I(\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i > \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_j) I(\delta_i = 1) / \hat{G}^2(y_i|\boldsymbol{x}_i)}{\sum_i \sum_j I(y_i < y_j) I(\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i \neq \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_j) I(\delta_i = 1) / \hat{G}^2(y_i|\boldsymbol{x}_i)}$$

where y is the minimum of the survival time and censoring time , $\delta = 1$ indicates that the survival time is smaller, and $\hat{G}(y_i|\boldsymbol{x}_i)$ is the within group Kaplan-Meier estimated conditional survivor function for the censoring time random variable. When the weighted c-index is asymptotically unbiased, its limiting value is $\Pr[\boldsymbol{\beta}^T \boldsymbol{x}_1 > \boldsymbol{\beta}^T \boldsymbol{x}_2 | T_2 > T_1, \boldsymbol{\beta}^T \boldsymbol{x}_1 \neq \boldsymbol{\beta}^T \boldsymbol{x}_2]$. It is shown in the appendix that if proportional hazards holds then this conditional probability is equal to the concordance probability.

The results shown in Table 1 demonstrate the accuracy and stability of the CPE. The bias in the CPE is small and is unaffected by the percent censoring, except when

the concordance probability is small. The small bias that occurs when the regression model is weak and censoring is high is a result of the bias in the Cox regression coefficients in these simulations. In contrast, the weighted c-index has a systemic bias that is an increasing function of the percent censoring. The standard error of the CPE increases as the percent censoring increases, and the strength of the model decreases. The estimated standard error of the CPE is close to its simulation standard error.

4. Prostate cancer example

Staging or risk classification for metastatic cancer is complicated due to the difficulty in measuring disseminated disease. For the metastatic prostate cancer population, an exploration of a risk classification system was undertaken based on a recently completed clinical trial of 1195 patients randomized 2:1 to either Abiraterone acetate or placebo (de Bono et al., 2011). In addition to the randomization, patients were stratified by: 1) the Eastern Cooperative Oncology Group performance status score, 2) the level of worst pain over the previous 24 hours on the Brief Pain Inventory - Short Form, 3) the number of previous chemotherapy regimens, and 4) the type of disease progression. For this analysis, only the 314 placebo patients with evaluable baseline markers were used to create a prognostic risk classification system.

Risk classification was based on baseline values of circulating tumor cells (CTC) and lactate dehydrogenase (LDH). These have been shown to be strong prognostic factors in this patient population (Scher et al., 2009). CTC is a blood-based assay that provides information on the accumulation of tumor cells in the peripheral blood

and LDH is a marker of disease burden. CTC was dichotomized, using the FDA approved classification, 4 or fewer cells/7.5 ml of blood and 5 or more cells/7.5 ml of blood. The dichotomization of LDH was based on its defined upper limit of normal, 250 U/L. The risk classification derived from these two factors was:

Low risk	CTC ≤ 5
Intermediate risk	CTC > 5 LDH ≤ 250
High risk	CTC > 5 LDH > 250

The Kaplan-Meier estimates of survival for the three risk groups are depicted in Figure 1. A proportional hazards model was generated using this risk classification and a test of the proportional hazards assumption developed by Grambsch and Therneau (1994) provided insufficient evidence that the proportional hazards assumption was violated (p= 0.11). From the results of the proportional hazards model, the estimated log relative risk of death was 0.759 (se=0.195) and 1.455 (se = 0.176) for the intermediate and high risk groups respectively, compared to the low risk group. This risk classification is clearly prognostic, but p-values alone, which were both <0.001, are insufficient to assess the strength of this classification. In general, p-values are sensitive to the sample size of the data, with a large number of failures producing a downward influence on the p-value. In order to ascertain the strength of this classification system in discriminating patient risk, the concordance probability estimate was computed.

The CPE accounting for the ties in the covariates developed by the three risk groups is equal to 0.741 (se=0.024). This result indicates that the simple three group staging system provides moderate discriminatory power, where a CPE above 0.80

demonstrates strong discrimination. If, however, the ties are included by substituting 1/2 for each pairwise tie in the calculation, then the CPE is equal 0.654 (se=0.015), a substantial decrease in the CPE value. This is due to the sizeable number of ties between patients classified to one of only three risk groups.

In addition to an overall effect, the proportional hazards specification can be used to estimate the model's discriminatory power between individual risk groups

$$\Pr[T_2 > T_1 | \boldsymbol{\beta}^T \boldsymbol{x}_1 = r_1, \boldsymbol{\beta}^T \boldsymbol{x}_2 = r_2] = [1 + \exp(r_2 - r_1)]^{-1}.$$

Using the log relative risk estimates, a patient in the low risk group has approximately 4:1 odds of surviving longer than a patient in the high risk group. The odds are approximately 2:1 that a patient in the low risk group survives longer than a patient in the intermediate risk group, and approximately the same odds for a patient in the intermediate risk group relative to a patient in the high risk group. This illustrates that although the overall measure of model strength is 0.741 (approximate 3:1 odds), the discriminatory strength between individual groups varies. The overall CPE is a weighted average of these individual measures.

5. Discussion

The interpretation of the CPE differs depending on whether ties in the risk groups are included or excluded. The CPE including ties provides a discrimination index for all patients, including subjects in the same clinical stage. The inclusion of ties penalizes model strength for subjects in the same risk group and therefore is affected by the distribution of the subjects across stages. The CPE excluding ties measures the model discriminatory power for patients in different clinical states, and is unaffected

by the distribution of patients across stage.

The most commonly applied measure of discrimination is the c-index. An advantage of the c-index is that it is not dependent on the proportional hazards model, and may be useful if alternative models (such as proportional odds or accelerated failure time models) are used to determine patient risk. We previously showed that the unweighted c-index had a bias that was an increasing function of the percent censored (Gönen and Heller, 2005). The inverse probability weighted c-index also has a positive bias, although not as large as the unweighted version. In application, the c-index employed on low risk cohorts where the number of events are small, will make a model appear stronger than it actually is. Uno et al. (2011) suggest that a partial weighted c-index $c(t)$, calculated to follow up time t, be computed to reduce the bias. It is unclear, however, how to choose t to balance the reduction in bias with the decrease in efficiency.

The CPE is useful if applied in conjunction with the proportional hazards model, and is virtually unaffected by independent censoring. If the proportional hazards model is applied to assess patient risk, and diagnostics are used to confirm the proportional assumptions, then the CPE is an accurate and robust measure of discrimination. `R` code to compute the CPE and its standard error, with ties included or ties excluded, is available in the `R` package `CPE`.

Appendix

The asymptotic distribution of the CPE.

To attain the asymptotic distribution of the CPE, it is approximated with the smooth statistic

$$\tilde{K}_n(\hat{\boldsymbol{\beta}}) = \frac{\frac{2}{n(n-1)} \sum\sum_{i<j} J_{ij} \left\{ \frac{\Phi(-\hat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{x}_{ji}/h)}{1+\exp(\hat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{x}_{ji})} + \frac{\Phi(-\hat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{x}_{ij}/h)}{1+\exp(\hat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{x}_{ij})} \right\}}{\frac{2}{n(n-1)} \sum\sum_{i<j} J_{ij} \{\Phi(-\hat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{x}_{ji}/h) + \Phi(-\hat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{x}_{ij}/h)\}}$$

where Φ represents the standard normal cumulative distribution function and $J_{ij} = I(\boldsymbol{x}_{ji} \neq \mathbf{0})$. The local Gaussian distribution function is introduced to create a statistic continuous in $\boldsymbol{\beta}$ for the Taylor expansion that follows. The bandwidth in the smoothed CPE was chosen as $h = 0.5\hat{\sigma} n^{-1/3}$, where $\hat{\sigma}$ is the estimated standard deviation of the linear combination $\hat{\boldsymbol{\beta}} \boldsymbol{x}_i$, computed for each subject. The rate $n^{-1/3}$ creates an asymptotic equivalence between the smooth and unsmooth CPE (Gönen and Heller, 2005).

The Taylor expansion of the smooth CPE produces

$$\tilde{K}_n(\hat{\boldsymbol{\beta}}) = \tilde{K}_n(\boldsymbol{\beta}) + \left(\frac{\partial \tilde{K}_n}{\partial \boldsymbol{\beta}} \right)^{\mathrm{T}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1)$$

The $\partial \tilde{K}_n / \partial \boldsymbol{\beta}$ is asymptotically constant. The component $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ has asymptotic mean zero, conditional on $\boldsymbol{x}$, and so is asymptotically independent of $\tilde{K}_n(\boldsymbol{\beta})$. Therefore the asymptotic variance of $\tilde{K}_n(\hat{\boldsymbol{\beta}})$ is

$$\mathrm{Var}[\tilde{K}_n(\hat{\boldsymbol{\beta}})] = \mathrm{Var}[\tilde{K}_n(\boldsymbol{\beta})] + \left(\frac{\partial \tilde{K}_n}{\partial \boldsymbol{\beta}} \right)^{\mathrm{T}} \mathrm{Var}[\hat{\boldsymbol{\beta}}] \left(\frac{\partial \tilde{K}_n}{\partial \boldsymbol{\beta}} \right) + o_p(1)$$

The individual components can be estimated as follows. The $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ is estimated from the second derivative of the partial likelihood.

To compute the partial derivative $\frac{\partial \tilde{K}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$, let

$$u_{ji2}(\boldsymbol{\beta}) = J_{ij}\Phi(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_{ji}/h), \qquad u_{ji1}(\boldsymbol{\beta}) = \frac{J_{ij}\Phi(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_{ji}/h)}{1+\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_{ji})}$$

and denote the derivative of each of these elements with respect to $\boldsymbol{\beta}$ as

$$u'_{ji2}(\boldsymbol{\beta}) = -J_{ij}(\boldsymbol{x}_{ji}/h)\phi(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_{ji}/h); \quad u'_{ji1}(\boldsymbol{\beta}) = \frac{(1+\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_{ji}))u'_{ji2} - u_{ji2}\boldsymbol{x}_{ji}\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_{ji})}{[1+\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_{ji})]^2}$$

where ϕ is the standard normal density function.

Then

$$\frac{\partial \tilde{K}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\sum\sum_{i<j}[u'_{ji1}(\boldsymbol{\beta}) + u'_{ij1}(\boldsymbol{\beta})]}{\sum\sum_{i<j}[u_{ji2}(\boldsymbol{\beta}) + u_{ij2}(\boldsymbol{\beta})]}$$

where we have used the identity $u'_{ji2}(\boldsymbol{\beta}) + u'_{ij2}(\boldsymbol{\beta}) = 0$.

To estimate the asymptotic variance of $\tilde{K}_n(\boldsymbol{\beta}_0)$, we note that

$$\tilde{K}_n(\boldsymbol{\beta}_0) = \frac{\kappa_1(\boldsymbol{\beta}_0)}{\kappa_2(\boldsymbol{\beta}_0)} \quad \text{and } \kappa_j \text{ are U-statistics of degree 2.}$$

From U-statistic theory

$$\sqrt{n}(\kappa_j(\boldsymbol{\beta}_0) - \mu_j(\boldsymbol{\beta}_0)) \sim N(0, v_j)$$

and

$$\frac{\kappa_1(\boldsymbol{\beta}_0)}{\kappa_2(\boldsymbol{\beta}_0)} \sim N\left(\frac{\mu_1(\boldsymbol{\beta}_0)}{\mu_2(\boldsymbol{\beta}_0)}, n^{-1}\boldsymbol{d}^{\mathrm{T}}(\boldsymbol{\beta}_0)V(\boldsymbol{\beta}_0)\boldsymbol{d}(\boldsymbol{\beta}_0)\right)$$

where

$$\boldsymbol{d}^{\mathrm{T}}(\boldsymbol{\beta}_0) = (\mu_2^{-1}(\boldsymbol{\beta}_0), \mu_1(\boldsymbol{\beta}_0)/\mu_2^2(\boldsymbol{\beta}_0)) \quad \text{and} \quad V(\boldsymbol{\beta}_0) = \mathrm{var}(\kappa_1(\boldsymbol{\beta}_0), \kappa_2^{\mathrm{T}}(\boldsymbol{\beta}_0)).$$

Estimating $\mu_j(\boldsymbol{\beta}_0)$ by $\kappa_j(\hat{\boldsymbol{\beta}})$, the estimated variance covariance matrix V is

$$v_{ab}(\hat{\boldsymbol{\beta}}) = \frac{4}{n(n-1)^2} \sum_i \sum_j \sum_{k \neq j} \left\{ [u_{ija}(\hat{\boldsymbol{\beta}}) + u_{jia}(\hat{\boldsymbol{\beta}})] - \kappa_a(\hat{\boldsymbol{\beta}}) \right\} \left\{ [u_{ikb}(\hat{\boldsymbol{\beta}}) + u_{kib}(\hat{\boldsymbol{\beta}})] - \kappa_b(\hat{\boldsymbol{\beta}}) \right\}$$

with $a, b = 1, 2$.

Combining these results provides the estimated asymptotic variance of the CPE.

Equality between the concordance probability and the limiting value of the c-index.

The concordance probability is

$$\mathrm{CP} = \Pr[T_2 > T_1 | \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_1 > \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_2]$$

and the limiting value of the c-index is

$$\mathcal{C} = \Pr[\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_1 > \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_2 | T_2 > T_1, \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_1 \neq \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_2].$$

First note that

$$\mathcal{C} = \mathrm{CP} \times \frac{\Pr[\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_1 > \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_2]}{\Pr[T_2 > T_1, \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_1 \neq \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_2]} \tag{A.1}$$

Assuming there are g risk group scores, the denominator in (A.1) is equal to

$$\begin{aligned}
&\Pr[T_2 > T_1, \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_1 \neq \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_2] \\
&= \sum_{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_1 = l_1}^{l_g} \sum_{\substack{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_2 = l_2 \\ l_2 \neq l_1}}^{l_g} \Pr[T_2 > T_1 | \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_1 = l_1, \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_2 = l_2] \times \Pr[\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_1 = l_1, \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_2 = l_2] \\
&= \sum_{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_1 > \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_2} \sum \Pr[\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_1 = l_1, \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_2 = l_2]
\end{aligned}$$

where the last equality follows from the proportional hazards identity

$$\Pr[T_2 > T_1 | \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_1 = l_1, \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_2 = l_2] = 1 - \Pr[T_2 > T_1 | \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_1 = l_2, \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_2 = l_1].$$

Substitution into the denominator of (A.1) proves the equality.

References

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

Cox, D.R. (1975). Partial likelihood. *Biometrika* **62**, 269-276.

de Bono, J. S., Logothetis, C. J., Molina, A., Fizazi, K., North, S., Chu, L., et al. (2011). Abiraterone and increased survival in metastatic prostate cancer. *New England Journal of Medicine* **364**, 1995-2005.

Edge, S. B., Byrd, D. R., Compton, C. C., Fritz, A. G., Green, F. L., Trotti, A. (Eds.) (2010). AJCC Cancer Staging Manual. Seventh edition. New York: Springer.

Gönen, M. and Heller, G. (2005). Concordance probability and discriminatory power in proportional hazards regression. *Biometrika* **92**, 965-970.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika*, **81**, 515-526.

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., Rosati, R. A. (1982). Evaluating the yield of medical tests. *Journal of the American Medical Association* **247**, 2543-2546.

Lin, D. Y., Wei, L. J., Ying, Z. (1993). Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. *Biometrika* **80**, 557-572.

Scher H. I., Jia, X., de Bono, J. S., Fleisher, M., Pienta, K. J., Raghavan, D., Heller, G. (2009). *Lancet Oncology* **10**, 233-239.

Uno, H., Tianxi, C., Pencina, M.J., D'Agostino, R. B., Wei, L. J. (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine* **30**, 1105-1117.

Table 1

Simulation results.

CP and limiting c-index	Censored proportion	Average IPCW c-index	Average CPE	Average CPE standard error	Simulation CPE standard deviation
0.600	0.000	0.603	0.605	0.027	0.025
0.600	0.251	0.615	0.606	0.031	0.027
0.600	0.501	0.624	0.609	0.038	0.033
0.600	0.751	0.636	0.620	0.054	0.041
0.701	0.000	0.702	0.703	0.024	0.022
0.701	0.247	0.713	0.703	0.028	0.026
0.701	0.498	0.722	0.703	0.034	0.032
0.701	0.744	0.730	0.702	0.048	0.045
0.799	0.000	0.799	0.800	0.020	0.019
0.799	0.250	0.807	0.800	0.023	0.022
0.799	0.498	0.820	0.801	0.028	0.026
0.799	0.751	0.830	0.796	0.042	0.039
0.900	0.000	0.900	0.900	0.017	0.013
0.900	0.250	0.906	0.901	0.019	0.015
0.900	0.497	0.913	0.901	0.024	0.018
0.900	0.746	0.927	0.900	0.035	0.026

CP = Concordance Probability; CPE = Concordance Probability Estimate;

IPCW = Inverse Probability Censoring Weight

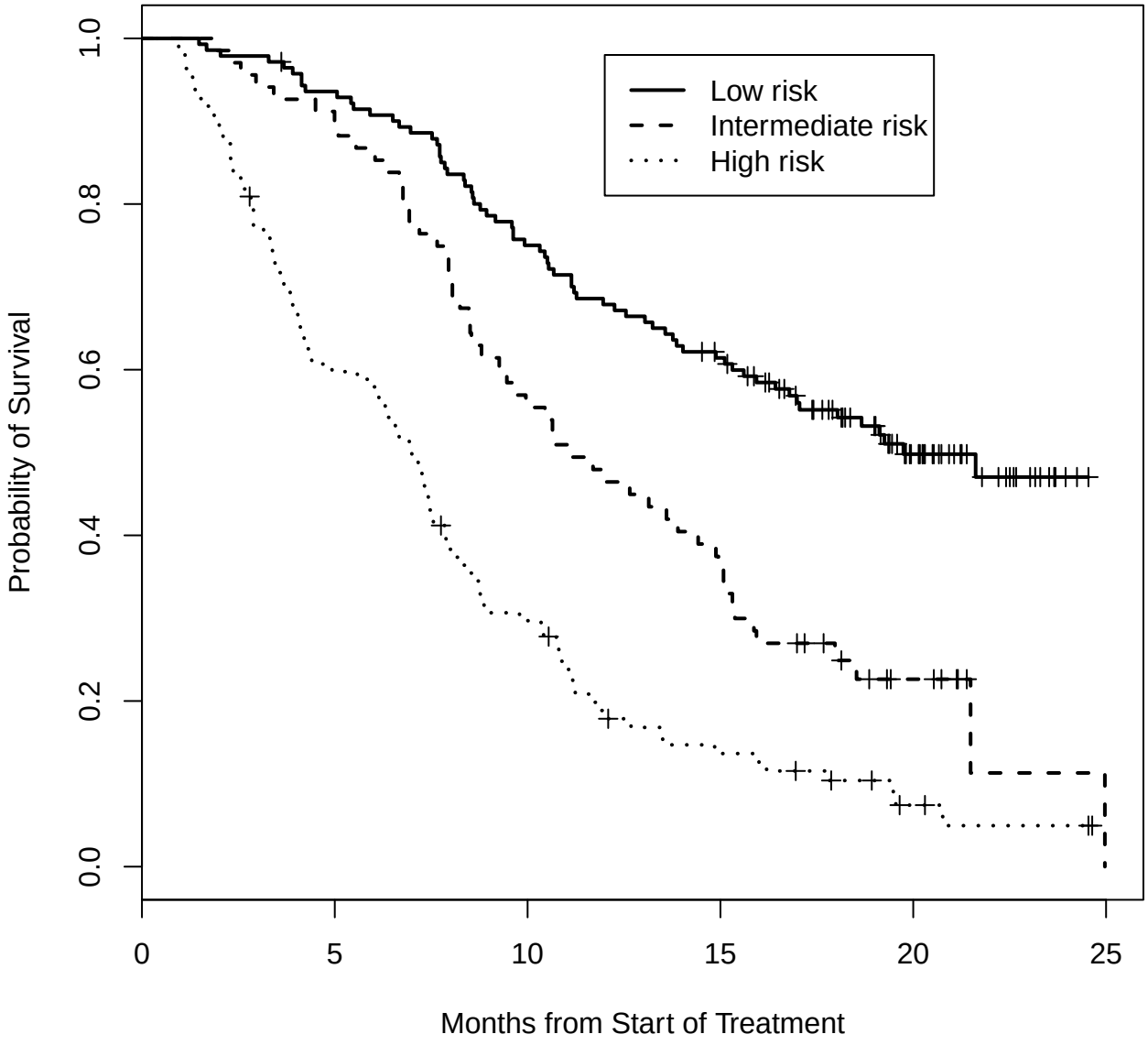

Figure 1. Kaplan-Meier estimates of survival based on clinical risk stage.