

Measuring the Impact of New Risk Factors Within Survival Models

Glenn Heller and Sean M. Devlin

Memorial Sloan Kettering Cancer Center, New York, USA

Abstract

Survival is poor for patients with metastatic cancer, and it is vital to examine new biomarkers that can improve patient prognostication and identify those who would benefit from more aggressive therapy. In metastatic prostate cancer, two new assays have become available: one that quantifies the number of cancer cells circulating in the peripheral blood, and the other a marker of the aggressiveness of the disease. It is critical to determine the magnitude of the effect of these biomarkers on the discrimination of a model-based risk score. To do so, most analysts frequently consider the discrimination of two separate survival models: one that includes both the new and standard factors and a second that includes the standard factors alone. However, this analysis is ultimately incorrect for many of the scale-transformation models ubiquitous in survival, as the reduced model is misspecified if the full model is specified correctly. To circumvent this issue, we developed a projection-based approach to estimate the impact of the two prostate cancer biomarkers. The results indicate that the new biomarkers

can influence model discrimination and justify their inclusion in the risk model; however, the hunt remains for an applicable model to risk-stratify patients with metastatic prostate cancer.

Keywords: Biomarkers; Concordance probability; Nested models; Projection; Prostate cancer

1 Introduction

Survival risk models are used to assess the risk of disease onset for an individual in the general population or the risk of disease progression or death for a patient already diagnosed with an illness. The search to identify new risk factors and biomarkers to improve survival risk models is ubiquitous in clinical research. For example in oncology, a large research effort is underway to evaluate patient risk as a function of molecular alterations within the tumor cell as assessed via a biopsy or peripheral blood. An important research question is whether this newly acquired cellular information impacts survival risk, after accounting for existing clinical measures and laboratory tests obtained during routine practice.

In this work, the interest is focused on the added value of two new blood-based biomarkers, circulating tumor cells (CTC) and serum testosterone, in men with metastatic prostate cancer. Contemporary assays, such as CTC and serum testosterone, are typically more expensive to administer than the conventional assays used to assess patient risk in this population. To justify their routine use within a risk model, evidence beyond association metrics with survival is required. In this paper, we

seek a second level of evidence, specifically examining how the incorporation of these factors into the risk model influences the concordance probability.

Data from a randomized clinical trial (Saad et al., 2015) of men treated on one trial arm with an androgen receptor targeted therapy, a key molecular target in the metastatic prostate cancer population, was used to evaluate the added value of CTC and serum testosterone. The primary endpoint of the clinical trial was overall survival, and the median survival time was only 31 months. This relatively short survival accentuates that accurate baseline risk determination is critical in the patient population. Historically, the prostate specific antigen (PSA) biomarker, the metastatic site where the cancer has spread (bone or soft tissue), and the ability of the patient to perform routine tasks (ECOG status), have been employed to assess patient risk in the metastatic prostate cancer population. For the current analysis, these conventional risk factors along with two new risk factors, circulating tumor cells (CTC) and serum testosterone, were incorporated into a proportional hazards risk model.

A summary of the results from 631 men with complete risk factor data is provided in Table 1. All the risk factors, including CTC and serum testosterone, were prognostic for survival. To assess the adequacy of the proportional hazards assumption necessary for the interpretation of the model results, a smoothed relationship between the scaled Schoenfeld residuals and time is plotted for each risk factor (Grambsch and Therneau, 1994). A nonconstant slope in these graphs indicates nonproportionality. The residual plots and the p-value generated from a null test statistic of a constant slope are provided in Supplemental Figure S1. To achieve the proportional hazards specification in this enhanced model, the continuous risk factors, PSA, testosterone, and CTC were

all transformed to a square root scale. In addition, CTC was transformed into two constructed variables based on whether the patient's CTC value exceeded 10. The diagnostic results of this model specification are consonant with the proportional hazards assumption across all risk factors.

Although all risk factors were found prognostic, with 248 events (deaths) in the data, it is not highly informative to determine that all the log relative risk parameters are non-zero. It is often the case that after this association analysis, a second-level analysis is performed to better understand the quality of the risk model. Common families of metrics used for this evaluation include calibration, discrimination, explained variation, and likelihood based. This work will focus on the concordance probability, a measure of discrimination that gauges the extent the model-based baseline risk score separates long-term survivors from short-term survivors. A high estimate of the concordance probability is an indication that the risk score can reliably determine patients who are at high risk and perhaps require more aggressive treatment.

The key question we address is how to evaluate the influence of the new factors, and more generally the influence of any subset of factors, on the concordance probability in a survival model. We will use the metastatic prostate cancer example throughout the paper to illustrate the limitations of a common approach to this assessment and to propose a new solution.

2 Concordance Probability

For survival models, a popular discrimination measure is the concordance probability. An early definition of the concordance probability, modified to reflect the analysis horizon τ , is

$$\Pr[S_t(\mathbf{X}_1, \mathbf{Z}_1) < S_t(\mathbf{X}_2, \mathbf{Z}_2) | T_1 < T_2, T_1 < \tau]$$

(Pencina and D'Agostino, 2004), where T is the survival time, \mathbf{X} represents a p -dimensional vector containing the conventional risk factors, \mathbf{Z} is a q -dimensional vector of new risk factors, and $S_t(\mathbf{X}, \mathbf{Z})$ denotes the survival function $\Pr(T > t | \mathbf{X}, \mathbf{Z})$. Throughout the paper, upper case letters represent random variables, lower case their realizations, and bold type indicate vectors. In addition, it is assumed that at least one conventional risk factor is continuous. The concordance probability is an expression of the concordance between the survival times and the predicted survival probability beyond a given time t . The concordance probability ranges between 0.5 and 1.0, with the value 1 indicating perfect concordance.

The survival function $S_t(\mathbf{X}, \mathbf{Z})$ is frequently model-based. The most common risk models applied in clinical research are scale transformation models (Cheng et al., 1995). The scale transformation model is represented as

$$m(T) = -\boldsymbol{\beta}^T \mathbf{X} - \boldsymbol{\gamma}^T \mathbf{Z} + \epsilon,$$

where $m(T)$ is an unknown monotone increasing function on $[0, \tau]$, with $m(0) = -\infty$ and $m(\tau) = M < \infty$. The linear combinations $(\boldsymbol{\beta}^T \mathbf{X}, \boldsymbol{\gamma}^T \mathbf{Z})$ are risk indices, and ϵ is a random error with known distribution independent of (\mathbf{X}, \mathbf{Z}) .

The scale transformation model may be rewritten as

$$g(S_T(\mathbf{X}, \mathbf{Z})) = m(T) + \beta^T \mathbf{X} + \gamma^T \mathbf{Z},$$

with $\epsilon = g(S_T(\mathbf{X}, \mathbf{Z}))$ and $g(\cdot)$ is a monotone decreasing function. This alternative representation follows from the probability integral transform of S , a uniform $(0, 1)$ random variable, leading to the distribution of ϵ through the specification of g . Popular examples of scale transformation models include (Dabrowska and Doksum, 1988):

1. Proportional hazards : $g(S) = \log\{-\log(S)\}$
2. Proportional odds : $g(S) = -\log\{S/(1 - S)\}$
3. Generalized probit: $g(S) = \bar{\Phi}^{-1}(S)$

where $\bar{\Phi}$ is the standard normal survival function.

Scale transformation models contain two simplifying properties. First, there is no loss in information by considering the index survival function $S_T(\beta^T \mathbf{X} + \gamma^T \mathbf{Z})$, and second, the survival function is a monotonic decreasing function of the risk index $\beta^T \mathbf{X} + \gamma^T \mathbf{Z}$. As a result, the concordance probability may be rewritten as

$$\kappa(\tau) = \Pr[\beta^T \mathbf{X}_1 + \gamma^T \mathbf{Z}_1 > \beta^T \mathbf{X}_2 + \gamma^T \mathbf{Z}_2 | T_1 < T_2, T_1 < \tau], \quad (1)$$

which is the consensus version of how the survival concordance probability is defined today (Uno et al., 2011).

The most widely used estimate of the concordance probability is the concordance index (Harrell et al., 1982, 1996). Uno et al. (2011) demonstrated that due to censoring, the concordance index (c-index) is not a consistent estimate of $\kappa(\tau)$ and

incorporated an inverse probability censoring weight correction. The weighted c-index is computed as

$$D_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{G}; \tau) = \frac{\sum_{i,j} I(\hat{\boldsymbol{\beta}}^T \mathbf{x}_i + \hat{\boldsymbol{\gamma}}^T \mathbf{z}_i > \hat{\boldsymbol{\beta}}^T \mathbf{x}_j + \hat{\boldsymbol{\gamma}}^T \mathbf{z}_j) \delta_i I(y_i < y_j, y_i < \tau) \{\hat{G}(y_i)\}^{-2}}{\sum_{i,j} \delta_i I(y_i < y_j, y_i < \tau) \{\hat{G}(y_i)\}^{-2}}, \quad (2)$$

where $\sum_{i,j}$ denotes a double sum taken over the n observations in the sample, $Y = \min(T, C)$, $\delta = I(T < C)$, and C represents the underlying random censoring time with survival function $G(\cdot)$. The parameter estimates $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ from the scale transformation model may be computed using Cheng et al. (1995), and the Kaplan-Meier estimate of the censoring survival distribution may be used to estimate $G(y)$ under the assumption that $C \perp\!\!\!\perp (T, \mathbf{X}, \mathbf{Z})$. The weighted c-index can be generalized to include conditional independence ($C \perp\!\!\!\perp T | \mathbf{X}, \mathbf{Z}$), by replacing the Kaplan-Meier estimate with a conditional survival estimate of $G(y | \mathbf{x}, \mathbf{z})$ (Gerds et al., 2013; Gerds, 2023). When the scale transformation model is properly specified, Uno et al. (2011) provide conditions where the weighted c-index statistic is consistent and asymptotic normal.

An alternative estimate of $\kappa(\tau)$ is the concordance probability estimate (Gönen and Heller, 2005; Zhang and Shao, 2018),

$$K_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\theta}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}; \tau), \hat{\pi}; \tau) = \frac{\sum_{i,j} I(\hat{\boldsymbol{\beta}}^T \mathbf{x}_i + \hat{\boldsymbol{\gamma}}^T \mathbf{z}_i > \hat{\boldsymbol{\beta}}^T \mathbf{x}_j + \hat{\boldsymbol{\gamma}}^T \mathbf{z}_j) \hat{\theta}_{ijij}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}; \tau)}{n(n-1)\hat{\pi}(\tau)}, \quad (3)$$

where $\pi(\tau) = \Pr[T_1 < T_2, T_1 < \tau]$ is the marginal precedence probability and

$$\theta_{1212}(\boldsymbol{\beta}, \boldsymbol{\gamma}; \tau) =$$

$$\Pr[T_1 < T_2, T_1 < \tau | \boldsymbol{\beta}^T \mathbf{X}_1 = \boldsymbol{\beta}^T \mathbf{x}_1, \boldsymbol{\beta}^T \mathbf{X}_2 = \boldsymbol{\beta}^T \mathbf{x}_2, \boldsymbol{\gamma}^T \mathbf{Z}_1 = \boldsymbol{\gamma}^T \mathbf{z}_1, \boldsymbol{\gamma}^T \mathbf{Z}_2 = \boldsymbol{\gamma}^T \mathbf{z}_2] \quad (4)$$

is the model precedence probability. Going forward, we will simplify the notation and write $\hat{\theta}$ for $\hat{\theta}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}; \tau)$ in the argument of the concordance probability estimate.

The concordance probability estimate (CPE) relies on the choice of the scale transformation model used to estimate the precedence probability. For the scale transformation models introduced earlier in this section, the precedence probabilities are

- Proportional hazards :

$$\theta_{1212}(\boldsymbol{\beta}, \boldsymbol{\gamma}; \tau) = \frac{1 - S_{\tau}(\mathbf{x}_1, \mathbf{z}_1)S_{\tau}(\mathbf{x}_2, \mathbf{z}_2)}{1 + \exp [\boldsymbol{\beta}^T \mathbf{x}_{12} + \boldsymbol{\gamma}^T \mathbf{z}_{12}]}$$

- Proportional odds :

$$\theta_{1212}(\boldsymbol{\beta}, \boldsymbol{\gamma}; \tau) = \exp [\boldsymbol{\beta}^T \mathbf{x}_{12} + \boldsymbol{\gamma}^T \mathbf{z}_{12}] \times \frac{\left\{ \boldsymbol{\beta}^T \mathbf{x}_{12} + \boldsymbol{\gamma}^T \mathbf{z}_{12} + \log \left(\frac{1 - S_{\tau}(\mathbf{x}_2, \mathbf{z}_2)}{1 - S_{\tau}(\mathbf{x}_1, \mathbf{z}_1)} \right) \right\} + (1 - \exp [\boldsymbol{\beta}^T \mathbf{x}_{12} + \boldsymbol{\gamma}^T \mathbf{z}_{12}]) (1 - S_{\tau}(\mathbf{x}_2, \mathbf{z}_2))}{(1 - \exp [\boldsymbol{\beta}^T \mathbf{x}_{12} + \boldsymbol{\gamma}^T \mathbf{z}_{12}])^2}$$

- Generalized probit:

$$\theta_{1212}(\boldsymbol{\beta}, \boldsymbol{\gamma}; \tau) = \Phi_{D,M} \left[\frac{\boldsymbol{\beta}^T \mathbf{x}_{12} + \boldsymbol{\gamma}^T \mathbf{z}_{12}}{\sqrt{2}}, \bar{\Phi} (S_{\tau}(\mathbf{x}_1, \mathbf{z}_1)) \right]$$

where $\mathbf{x}_{12} = \mathbf{x}_1 - \mathbf{x}_2$, $\mathbf{z}_{12} = \mathbf{z}_1 - \mathbf{z}_2$, and $\Phi_{D,M}$ represents a bivariate standard normal distribution function, where the first component is the difference of two independent standard normals and the second component is a marginal standard normal. The proportional hazards and proportional odds models are specific cases of the Pareto family of models. However, as studied in Brentnall and Cuzick (2018), the precedence probability in the general Pareto model cannot be derived analytically.

The model precedence probability estimate is a function of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{S}_{\tau}(\mathbf{X}, \mathbf{Z}))$. Estimates of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ may be computed using Cheng et al. (1995). With g assumed

known, estimation of the model survival function is completed with the estimation of $m(\tau)$, which at the single point τ , is attained as the solution to the estimating equation (Cheng et al. 1997)

$$\sum_i \left\{ \frac{I(y_i \geq \tau)}{\hat{G}(\tau)} - g^{-1}[m(\tau) + \hat{\beta}^T \mathbf{x}_i + \hat{\gamma}^T \mathbf{z}_i] \right\} = 0.$$

The CPE in (3) is a consistent, asymptotic normal estimate of $\kappa(\tau)$ when the scale transformation model is properly specified (Devlin and Heller, 2021). Its consistency is apparent by applying Bayes Theorem in (1) to obtain

$$\kappa(\tau) = [\pi(\tau)]^{-1} \int \int \int \int_{\beta^T \mathbf{x}_{12} + \gamma^T \mathbf{z}_{12} > 0} \theta_{1212}(\boldsymbol{\beta}, \boldsymbol{\gamma}; \tau) dF(\boldsymbol{\beta}^T \mathbf{x}_1, \boldsymbol{\beta}^T \mathbf{x}_2, \boldsymbol{\gamma}^T \mathbf{z}_1, \boldsymbol{\gamma}^T \mathbf{z}_2). \quad (5)$$

The CPE does not require an inverse probability censoring weight, making it less sensitive to censoring in the right tail than the weighted c-index, and is more efficient than the weighted c-index under the correct model specification. Thus, the CPE is advantageous to apply under proper model specification. However, if the enhanced model is incorrectly specified, the weighted c-index combined with the partial rank estimate (Section 3.3), is more robust to model misspecification and has greater efficiency than the CPE. These assertions are supported by the simulations in Section 4.

3 Impact of New Factors on the Concordance Probability

3.1 Nested Models Approach

To determine the impact of a new set of factors on the concordance probability, the current practice is to use the estimates from nested scale transformation models

$$m(T) = -\boldsymbol{\beta}^T \mathbf{X} - \boldsymbol{\gamma}^T \mathbf{Z} + \epsilon \quad (6)$$

$$m^*(T) = -\boldsymbol{\beta}^{*T} \mathbf{X} + \epsilon^* \quad (7)$$

and calculate the difference in the estimated concordance probabilities computed from the enhanced model (6) and the reduced model (7). The expectation is that the addition of new factors to the model will improve the understanding of patient risk.

For the metastatic prostate cancer data, the results from the enhanced model were presented in Section 1. The results for the reduced proportional hazards model, based on the conventional risk factors, PSA, visceral disease, and ECOG status, are presented in Table 2. The conventional risk factors remain prognostic; however, the proportional hazards assumption does not hold for visceral disease, as illustrated in Supplemental Figure S2. This scenario, where the proportional hazards assumption is satisfied in the investigator-constructed enhanced model, but violated in the reduced model is unsurprising, since proportional hazards models are, in general, non-nested (Hougaard, 1986; Fine, 2002). Thus, a limitation of comparing estimates of the concordance probabilities from nested scale transformation models is the structural misspecification of the reduced model due to the non-nesting property of proportional hazards models. This misspecification will affect the estimated concordance probability from the reduced model using either the weighted c-index or CPE and, ultimately, the interpretation of the impact of circulating tumor cells and serum testosterone on this discrimination measure.

The non-nesting argument for the proportional hazards assumption may be generalized

to the family of scale transformation models. If the heterogeneity due to \mathbf{Z} in the enhanced scale transformation model (6) is ignored by incorporating $\gamma^T \mathbf{Z}$ into the error term, the reduced model is written as (7), where ϵ^* is a location-scale transformation of $\epsilon - \gamma^T \mathbf{Z}$. In general, the distribution of ϵ^* is not the same as ϵ , and hence the scale transformation family models are nonnested. The generalized probit model is an exception if the distribution of (\mathbf{X}, \mathbf{Z}) is multivariate normal.

A measure for the impact of the new factors to the concordance probability, based on the nested models in (6) and (7), is

$$\Pr[\beta^T \mathbf{X}_1 + \gamma^T \mathbf{Z}_1 > \beta^T \mathbf{X}_2 + \gamma^T \mathbf{Z}_2 | T_1 < T_2, T_1 < \tau] - \Pr[\beta^{*T} \mathbf{X}_1 > \beta^{*T} \mathbf{X}_2 | T_1 < T_2, T_1 < \tau].$$

This metric is a function of the relationship between the new risk factors \mathbf{Z} and survival time T in the presence of the conventional risk factors \mathbf{X} and the degree of misspecification of the reduced model. It is clearly undesirable for the influence of the new factors on the concordance probability to be a function of the level of misspecification. Thus, we will not proceed further with this nested model approach.

In the next two sections, alternative approaches that do not rely on the misspecified reduced model but instead use a projection approach, are provided to compute the impact of the new factors on the concordance probability. In Section 3.2, the methodology is applied when the error distribution for ϵ from a scale transformation model is known, and Section 3.3 considers the case that the error distribution from this model is unknown. Estimation for the concordance probability and inference for the impact parameter are carried out separately depending on the error distribution information.

3.2 Error Distribution Known

To develop a concordance probability measure that ignores the heterogeneity due to the new factor risk index $\boldsymbol{\gamma}^T \mathbf{Z}$, the enhanced model precedence probability (4) is projected onto the two-dimensional space spanned by the random variables $(\boldsymbol{\beta}^T \mathbf{X}_1, \boldsymbol{\beta}^T \mathbf{X}_2)$,

$$\theta_{12}^{[P]}(\boldsymbol{\beta}; \tau) = E [\theta_{1212}(\boldsymbol{\beta}, \boldsymbol{\gamma}; \tau) | \boldsymbol{\beta}^T \mathbf{X}_1, \boldsymbol{\beta}^T \mathbf{X}_2], \quad (8)$$

which as a direct result of iterated expectation

$$\theta_{12}^{[P]}(\boldsymbol{\beta}; \tau) = \Pr[T_1 < T_2, T_1 < \tau | \boldsymbol{\beta}^T \mathbf{X}_1, \boldsymbol{\beta}^T \mathbf{X}_2].$$

Through this projected precedence probability, the concordance probability, ignoring the new factors in the enhanced model, is defined in an analogous way to (5) as

$$\kappa^{[P]}(\tau) = [\pi(\tau)]^{-1} \int \int_{\boldsymbol{\beta}^T \mathbf{x}_{12} > 0} \theta_{12}^{[P]}(\boldsymbol{\beta}; \tau) dF(\boldsymbol{\beta}^T \mathbf{x}_1, \boldsymbol{\beta}^T \mathbf{x}_2). \quad (9)$$

which equals

$$\Pr[\boldsymbol{\beta}^T \mathbf{X}_1 > \boldsymbol{\beta}^T \mathbf{X}_2 | T_1 < T_2, T_1 < \tau]. \quad (10)$$

The concordance probability from the enhanced model $\kappa(\tau)$, and the concordance probability for the conventional factor risk index alone $\kappa^{[P]}(\tau)$, may each be interpreted as projections. $\kappa(\tau)$ is proportional to the projection of the random variable $I(T_1 < T_2, T_1 < \tau)$ onto the half-space $\boldsymbol{\beta}^T \mathbf{X}_{12} + \boldsymbol{\gamma}^T \mathbf{Z}_{12} > 0$ and $\kappa^{[P]}(\tau)$ is proportional to the projection of this random variable onto the half-space $\boldsymbol{\beta}^T \mathbf{X}_{12} > 0$. The difference in the concordance probabilities leads to a measure of the impact of the new factors in the enhanced model

$$\xi(\tau) = \kappa(\tau) - \kappa^{[P]}(\tau). \quad (11)$$

The efficiency of the CPE relative to the weighted c-index when the error distribution in (6) is known, suggests that the estimated impact parameter be computed as

$$\hat{\xi}(\tau) = K_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\theta}, \hat{\pi}; \tau) - K_n^{[P]}(\hat{\boldsymbol{\beta}}, \hat{\theta}^{[P]}, \hat{\pi}; \tau),$$

where $K_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\theta}, \hat{\pi}; \tau)$ is expressed in (3) and

$$K_n^{[P]}(\hat{\boldsymbol{\beta}}, \hat{\theta}^{[P]}, \hat{\pi}; \tau) = \frac{\sum_{i,j} I[\hat{\boldsymbol{\beta}}^T \mathbf{x}_i > \hat{\boldsymbol{\beta}}^T \mathbf{x}_j] \hat{\theta}_{ij}^{[P]}(\hat{\boldsymbol{\beta}}; \tau)}{n(n-1) \hat{\pi}(\tau)}.$$

The projected precedence probability $\theta_{12}^{[P]}(\boldsymbol{\beta}; \tau)$, a conditional expectation (8), is estimated using kernel smoothing

$$\hat{\theta}_{12}^{[P]}(\hat{\boldsymbol{\beta}}; \tau) = \frac{\sum_{k,l} \hat{\theta}_{12kl}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}; \tau) \phi_h(\hat{\boldsymbol{\beta}}^T \mathbf{x}_1, \hat{\boldsymbol{\beta}}^T \mathbf{x}_k) \phi_h(\hat{\boldsymbol{\beta}}^T \mathbf{x}_2, \hat{\boldsymbol{\beta}}^T \mathbf{x}_l)}{\sum_{k,l} \phi_h(\hat{\boldsymbol{\beta}}^T \mathbf{x}_1, \hat{\boldsymbol{\beta}}^T \mathbf{x}_k) \phi_h(\hat{\boldsymbol{\beta}}^T \mathbf{x}_2, \hat{\boldsymbol{\beta}}^T \mathbf{x}_l)},$$

where $\phi_h(u, v)$ is a kernel function with bandwidth h . The justification for the projected precedence probability estimate is provided in Theorem 1.

Theorem 1:

Let $\phi_h(u, v)$ represent a kernel function, symmetric about zero, with bandwidth h . Assume as $n \rightarrow \infty$, $h \rightarrow 0$ and $nh^2 \rightarrow \infty$. Then

$$\hat{\theta}_{12}^{[P]}(\hat{\boldsymbol{\beta}}; \tau) \xrightarrow{p} \Pr(T_1 < T_2, T_1 < \tau | \boldsymbol{\beta}^T \mathbf{X}_1, \boldsymbol{\beta}^T \mathbf{X}_2)$$

Theorem 1 is derived in the appendix and follows the conventional asymptotic result using a Nadaraya-Watson estimator (Simonoff, 1996).

The asymptotic distribution of $\hat{\xi}(\tau)$ is developed to obtain its asymptotic variance and to produce a 95% confidence interval for $\xi(\tau)$. These inferential measures will be applied in Section 5 to complete the prostate cancer data analysis. The asymptotic distribution for the CPE impact estimate is provided in Theorem 2 and is derived in the appendix.

Theorem 2:

Assume the scale transformation model based on the covariates (\mathbf{x}, \mathbf{z}) is properly specified, and the new factors are associated with survival time ($\gamma \neq 0$). Then

$$n^{1/2}[K_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}, \hat{\pi}; \tau) - K_n^{[P]}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}^{[P]}, \hat{\pi}; \tau) - \xi(\tau)] \xrightarrow{D} N(0, V_K). \quad (12)$$

The result is derived by decomposing (12) into components and demonstrating the asymptotic normality of each component. An analytic estimate of the asymptotic variance V_K via the sum of these components is complex, and a bootstrap estimate of the asymptotic variance and confidence interval will be applied (Kosorok et al., 2004).

3.3 Error Distribution Unknown

In the previous section, the impact of the new factors on the concordance probability was developed under the assumption that the error distribution was known. The working premise is that the analyst has carefully evaluated the data and chosen from the family of enhanced scale transformation models in Section 2. However, there will be times when even a close approximation will not be suitable.

For these data, estimates of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ may be derived, without specifying an error distribution for ϵ , by using the partial rank methodology (Khan and Tamer, 2007). The partial rank estimates (PRE) of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ are obtained from the objective function

$$J_n(\boldsymbol{\eta}, \boldsymbol{\gamma}) = \arg \max_{(\boldsymbol{\eta}, \boldsymbol{\gamma})} [n(n-1)]^{-1} \sum_i \sum_j \delta_j I[y_i > y_j] I[\boldsymbol{\beta}^T \mathbf{x}_i + \boldsymbol{\gamma}^T \mathbf{z}_i < \boldsymbol{\beta}^T \mathbf{x}_j + \boldsymbol{\gamma}^T \mathbf{z}_j],$$

where to resolve an identifiability issue in this maximization, the first component of $\boldsymbol{\beta}$ is set to one, and $\hat{\boldsymbol{\beta}} = (1, \hat{\boldsymbol{\eta}}^T)^T$. Khan and Tamer (2007) prove that, under

regularity conditions, which includes that there is at least one continuous covariate with a non-zero coefficient, $(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}})$ are asymptotically normal and $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ are consistent estimates of the rescaled parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ in (6). However, the discontinuity in the objective function can result in instability in this estimation process. To stabilize the algorithm, a smooth version of the objective function is employed

$$\tilde{J}_n(\boldsymbol{\eta}, \boldsymbol{\gamma}) = \arg \max_{(\boldsymbol{\eta}, \boldsymbol{\gamma})} [n(n-1)]^{-1} \sum_i \sum_j \delta_j I[y_i > y_j] \Phi \left(\frac{\boldsymbol{\beta}^T (\mathbf{x}_j - \mathbf{x}_i) + \boldsymbol{\gamma}^T (\mathbf{z}_j - \mathbf{z}_i)}{g} \right),$$

where $\Phi(\cdot)$ represents a local normal distribution function and g is its bandwidth which converges to 0 as n increases (Song et al., 2007).

With the coefficient estimates $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$, the impact of the new factors on the weighted c-index may be estimated by

$$D_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{G}; \tau) - D_n^{[P]}(\hat{\boldsymbol{\beta}}, \hat{G}; \tau)$$

where $D_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{G}; \tau)$ is expressed in (2), and from (10),

$$D_n^{[P]}(\hat{\boldsymbol{\beta}}, \hat{G}; \tau) = \frac{\sum_{i,j} I[\hat{\boldsymbol{\beta}}^T \mathbf{x}_i > \hat{\boldsymbol{\beta}}^T \mathbf{x}_j] \delta_i I[y_i < y_j, y_i < \tau] \{\hat{G}(y_i)\}^{-2}}{\sum_{i,j} \delta_i I[y_i < y_j, y_i < \tau] \{\hat{G}(y_i)\}^{-2}},$$

is a consistent estimate of the projected concordance probability $\kappa^{[P]}(\tau)$. In contrast to Section 3.2, the weighted c-index is employed because the model precedence probability cannot be computed without specifying the error distribution. The use of the weighted c-index simplifies the calculation of the impact estimate, but at the cost of including inverse probability censoring weights. The asymptotic distribution for the estimated impact, using the weighted c-index, is provided in Theorem 3.

Theorem 3:

$$n^{1/2}[D_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{G}; \tau) - D_n^{[P]}(\hat{\boldsymbol{\beta}}, \hat{G}; \tau) - \xi(\tau)] \xrightarrow{D} N(0, V_D). \quad (13)$$

This result follows directly by noting that

$$D_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{G}; \tau) - D_n^{[PL]}(\hat{\boldsymbol{\beta}}, \hat{G}; \tau) = \frac{\sum_{i,j} \left[I(\hat{\boldsymbol{\beta}}^T \mathbf{x}_i + \hat{\boldsymbol{\gamma}}^T \mathbf{z}_i > \hat{\boldsymbol{\beta}}^T \mathbf{x}_j + \hat{\boldsymbol{\gamma}}^T \mathbf{z}_j) - I(\hat{\boldsymbol{\beta}}^T \mathbf{x}_i > \hat{\boldsymbol{\beta}}^T \mathbf{x}_j) \right] \delta_i I(y_i < y_j, y_i < \tau) \{\hat{G}(y_i)\}^{-2}}{\sum_{i,j} \delta_i I(y_i < y_j, y_i < \tau) \{\hat{G}(y_i)\}^{-2}}$$

and applying the asymptotic normality derivation in Uno et al. (2011) to the statistic above.

4 Simulation Studies

Simulations were performed to estimate the adequacy of the projection estimate of the concordance probability and the resultant impact estimate. The three approaches detailed in this work were considered.

1. The enhanced model coefficients were estimated with the partial likelihood, and the concordance probabilities were computed with the CPE. [PL/CPE]
2. The enhanced model coefficients were estimated with the partial likelihood, and the concordance probabilities were estimated with the weighted c-index. [PL/wCI]
3. The enhanced model coefficients were estimated using the partial rank method, and the concordance probabilities were computed with the weighted c-index. [PR/wCI]

The abbreviations in square brackets will be used freely throughout this section to denote the three approaches.

The first set of simulations evaluated the estimates under proportional hazards, while the second evaluated the estimates under non-proportional hazards. Across all simulations, two conventional factors, X_1 and X_2 , and two new factors under evaluation, Z_1 and Z_2 , were generated as independent normal random variables with mean 0 and variance 1. All four factors generated the underlying survival times for the enhanced model, and the censoring times were generated independently of all factors. Impact parameters of 0.025, 0.05, and 0.10 for the concordance probability were evaluated. Across all scenarios, the concordance probability of the enhanced model was 0.70.

For the proportional hazards simulations, survival times were generated using the regression model $t_i = \exp\{-(\beta_1 x_{i1} + \beta_2 x_{i2} + \gamma_1 z_{i1} + \gamma_2 z_{i2})\} \times \epsilon_i$, where ϵ_i were independent exponential random variables with a scale parameter equal to 1. To achieve an impact $\xi(\tau)$ of 0.025 on the concordance probability from the enhanced model, the following parameters were selected: $\beta_1 = 0.718$, $\beta_2 = 0.15$, $\gamma_1 = 0.346$, $\gamma_2 = 0.15$, and using $\tau = 1.18$. Censoring times were generated from a Uniform(0, b), where b was selected to be 1.58 and 4.75 to achieve an average censoring rate of 50% and 25%, respectively. The parameters were also estimated when the censoring rate was 0%, and all failure times were observed. To achieve an impact $\xi(\tau)$ of 0.05, the following parameters were selected: $\beta_1 = 0.624$, $\beta_2 = 0.15$, $\gamma_1 = 0.505$, $\gamma_2 = 0.15$. Lastly, for an impact of 0.10, the parameters were $\beta_1 = 0.408$, $\beta_2 = 0.15$, $\gamma_1 = 0.684$, and $\gamma_2 = 0.15$. The value of τ and the two censoring distributions remained the same.

Under non-proportional hazards, survival times were generated from a conditional proportional hazards model with a gamma frailty, $t_i = \exp\{-(\beta_1 x_{i1} + \beta_2 x_{i2} + \gamma_1 z_{i1} + \gamma_2 z_{i2} + \log w_i)\} \times \epsilon_i$, where w_i are independent gamma random variables with shape

and scale parameters equal to 0.25, and the ϵ_i were independent exponential random variables with a scale parameter of 1. To achieve an impact of 0.025 in the concordance probability, the following parameters were selected: $\beta_1 = 1.741$, $\beta_2 = 0.15$, $\gamma_1 = 0.887$, and $\gamma_2 = 0.15$, along with $\tau = 6$. Censoring times were again generated from $\text{Uniform}(0, b)$, where the values of 13 and 310 for b achieved an average censoring rate of 50% and 25%, respectively, in addition to the scenario when the rate was 0%. For an impact of 0.05, the parameters were $\beta_1 = 1.567$, $\beta_2 = 0.15$, $\gamma_1 = 1.301$, and $\gamma_2 = 0.15$. For an impact of 0.10, the parameters were $\beta_1 = 1.061$, $\beta_2 = 0.15$, $\gamma_1 = 1.754$, and $\gamma_2 = 0.15$. The values of τ and b remained the same.

For each iteration of the data-generating process, the impact of the new risk factors $\xi(\tau)$ on the concordance probability from the enhanced model, was estimated using the CPE projection framework via the estimates $K_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\pi}; \tau)$ and $K_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \hat{\pi}; \tau)$ and with the weighted c-index using $D_n(\hat{\beta}, \hat{\gamma}, \hat{G}; \tau)$ and $D_n^{[P]}(\hat{\beta}, \hat{G}; \tau)$ using the proportional hazards (PH) model estimates or the partial rank estimates (PR). Across 2,000 iterations, these approaches were compared based on the average bias and the relative efficiency using the root mean square error (rMSE) of the CPE with respect to the weighted c-index using either PH or PR. In addition, for each iteration of the data-generating process, the respective standard errors for each concordance measure were estimated using 50 bootstrap samples. Using these standard error estimates, the ratio of the estimated standard error to the simulation standard error was calculated along the 95% coverage probability. The sample size for all simulations was 300. The impact and projected concordance probability parameters were determined by simulation from the average of 2,000 iterations, with each iteration containing 2,000 uncensored observations.

Results for the proportional hazards simulations are presented in Tables 3, 4, and 5 when the impact parameter in the concordance probability was 0.025, 0.05, and 0.10, respectively, based on the two new risk factors. For the three approaches, the average biases for the projection and impact estimates were minimal across scenarios. Using the bootstrap-based standard error estimates, all methods provided coverage near the 95% nominal level for both parameters, and the ratios of the estimated standard error to the simulation standard error were aligned close to 1. Under correct specification of the underlying model, the PL/CPE had relative efficacy gains of 10-31% over the PL/wCI and the PR/wCI, where the gains depended on the censoring rate and whether the projection concordance probability or impact parameter was being estimated.

Tables 6, 7, and 8 provide the simulation results under the non-proportional hazards scenario when the impact on concordance was 0.025, 0.05, and 0.10, respectively. Due to the underlying model misspecification, the PL/CPE approach produced biased estimates for both the projection concordance and impact parameters, along with poor coverage. Importantly, the test for the proportional hazards assumption (Grambsch and Therneau, 1994) was rejected 100% when there was no censoring and at least 93% of the time across all scenarios (Supplemental Table, S1). This reiterates the need to interrogate any modeling assumptions before selecting the methodological approach to estimate the projection and impact parameters. While the average bias was small for the PL/wCI estimates, the simulation standard errors were misaligned with the estimated bootstrap standard errors in some scenarios. These situations led to inefficiencies and weakened coverage relative to the PR/wCI estimates.

These findings imply that in cases where the data closely align with the proportional

hazards assumption, the partial likelihood/CPE projection approach outperforms in estimating the influence of new risk factors on the concordance probability. This superiority is attributed to the generation of smaller standard error estimates and, consequently, the formation of more tightly bound 95% confidence intervals. However, when the data are not well approximated by proportional hazards, or more generally, the models in Section 2, the partial rank estimate/weighted c-index is the preferred approach.

5 Prostate Cancer Example - Conclusion

Six hundred and thirty one men with complete risk factor data were treated with an androgen receptor targeted therapy as part of a randomized clinical trial in metastatic prostate cancer. The objective was to use the survival and risk factor data to evaluate the impact of the two new biomarkers on the concordance probability. The rationale for the inclusion of the new biomarkers, circulating tumor cells and serum testosterone, are distinct. Circulating tumor cells provide a blood-based quantification of tumor burden with high levels a direct measure of advanced disease. In contrast, low values of serum testosterone in the presence of metastatic disease is an indirect indication that the prostate cancer is aggressive.

The maximum event time in this cohort was 32.7 months, which led to the choice $\tau = 30$ months for the follow-up time used in the concordance probability analysis. The CPE for the enhanced model was 0.704 and the 95% confidence interval for the concordance probability parameter (1) was (0.679 – 0.729). Applying the

projection method in Section 3.2, the CPE estimated impact parameter, computed from $K_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\pi}; \tau)$ and $K_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \hat{\pi}; \tau)$, was 0.088, and the 95% confidence interval for the impact parameter (11) was (0.060 – 0.115). The weighted c-index for the enhanced model using the partial rank estimate was 0.735, with its attendant 95% confidence interval (0.695 – 0.775). The weighted c-index produced a comparable impact estimate, computed from $D_n(\hat{\beta}, \hat{\gamma}, \hat{G}; \tau)$ and $D_n^{[P]}(\hat{\beta}, \hat{G}; \tau)$, equal to 0.095, and the 95% confidence interval for the impact parameter was (< 0.01 – 0.190). As anticipated, the confidence interval using the partial rank estimate was wider than the CPE-based interval, which results from the improved efficiency of the CPE when the data are well approximated by proportional hazards.

Historically, PSA has been one of the foremost prognostic serum biomarkers employed in prostate cancer research. When subtracting the standard CPEs from two nested incorrectly specified proportional hazards models, namely (i) ECOG status and visceral disease presence, and (ii) PSA, ECOG status, and visceral disease presence, the impact of removing PSA from the conventional risk model resulted in a CPE reduction of -0.019 (95% CI: $-0.046, 0.009$). This suggests either no impact or a marginal decrease in PSA’s prognostic impact on the CPE among the conventional risk factors, and contradicts the conventional wisdom in the field. The proportionality of the enhanced model including CTC and serum testosterone, however, may be leveraged to employ the projection method to gauge the impact of PSA on the CPE among conventional factors. For this purpose, the impact of CTC and testosterone on the enhanced model was subtracted from the impact of PSA, CTC, and testosterone on the enhanced model. The difference was 0.031 (95% CI: $0.008, 0.054$), supporting the positive influence of PSA on the CPE within the conventional risk factors.

These results provide unequivocal support for incorporating PSA, circulating tumor cells and testosterone into the enhanced risk model. However, the estimates of the concordance probability from the enhanced model containing all the risk factors (0.704 and 0.735) indicate that further research is needed to generate new risk factors in this population before the model may be applied to assist in clinical decisions.

6 Discussion

In biology and medicine, the discovery of novel biological factors is often followed by an assessment of their association with one or a set of clinical outcomes, a key step as a discovery goes from the bench to the bedside. If an association is established, subsequent empirical research is needed to evaluate the impact of these novel factors relative to known risk factors in the patient population. This step, which often requires a model to link the risk factors to the clinical outcome, is useful in determining where the discovery lies on the spectrum between redundant and orthogonal risk factor information.

The value of a risk model may be partly summarized using measures such as calibration, discrimination, likelihood-based, and explained variation. In this work, we focused on the concordance probability, a discrimination metric that synthesizes well when applied to scale transformation models. Naively, one approach to evaluate the influence of the new risk factors is to compute the concordance probability in a model containing both conventional risk factors and new risk factors and compare it to the concordance probability using the conventional risk factors alone. The issue

with this frequently utilized approach stems from the fact that scale transformation models, in general, do not exhibit nesting. This lack of nesting may subsequently affect estimation of the influence of new factors on the concordance probability.

To ascertain the impact of the new risk factors on the concordance probability, a new approach using projection methodology was developed with the concordance probability estimate (CPE) and the weighted c-index. This approach does not apply a reduced model, enabling the analyst to focus on the discriminative aspects of the newly introduced risk factors within the enhanced survival model. The choice of concordance estimates was dependent on whether the analyst was confident that the data fit one of the specified models in Section 2. The most commonly applied model in this family is the proportional hazards model. One benefit of identifying the error distribution (Section 3.2) is greater efficiency of the impact statistic through the use of the CPE. An additional benefit is the interpretability of the parameters (β, γ) in the scale transformation model. As noted in Dabrowska and Doksum (1988), the interpretation of the parameters for the models considered in Section 2 are independent of the transformation $m(\cdot)$.

The CPE approach assumes that the analyst has a good understanding of their data and can identify an enhanced scale transformation model to approximate this data. In practice, this may require covariate stratification, truncation of the follow up time, or the transformation of covariates. In the prostate data, the effect of circulating tumor cells (CTCs) was bifurcated based on whether the baseline value was less than or greater than ten CTCs to approximate a proportional hazards specification. There will be times, however, when one cannot approximate a properly specified enhanced scale transformation model. This includes the special case where it is the

reduced model that can be properly specified. In these cases, it is recommended that the analyst use the partial rank methodology to estimate the coefficients from the enhanced model and apply the weighted c-index along with its projection to evaluate the impact of the new covariates. One alternative to the partial rank estimate in this scenario, is to find a superset of factors that are properly specified by a scale transformation model, and evaluate the two subsets of risk factors using the projection method. This was the method we used to determine the impact of PSA among the conventional risk factors. A second method suggested by a reviewer, is to consider nonparametric maximum likelihood estimation (Zeng and Lin, 2007), which requires joint estimation of the infinite dimensional parameter $m(t)$ and the finite dimensional coefficient parameters (β, γ) .

For the prostate cancer data in this paper, the concordance probability estimate, measured on a scale between 0.5 and 1.0, was only 0.70 in the enhanced model, an indication that further research is needed to find new factors that discriminate between longer-term survivors and short-term survivors. The analysis also demonstrated that the new biomarkers CTC and serum testosterone were highly influential in discriminating patient risk. The benefit of CTC alone can be further magnified using the proposed projection methodology. Ignoring the heterogeneity due to CTC in the enhanced risk model, produced a much smaller CPE equal to 0.62. Alternatively, ignoring all factors but CTC in the risk model, produced a concordance probability estimate equal to 0.67, surprisingly close to the enhanced model CPE. Thus, although the CTC assay costs more than the combined costs of the PSA and serum testosterone assay, there is a benefit to its inclusion in the risk model. Further research evaluating a cost-benefit analysis for CTC and overall survival would provide additional useful

information.

The impact measure was derived from a scale transformation survival model. The benefit of the scale transformation model is the reduction in dimensionality in the risk assessment through a single risk index. The application of a single index model is justified as an approximation to a more complex risk function (Hall, 1989). Alternative approaches to understanding individual risk, which explicitly encompasses greater risk complexity and dimensionality, include random forest, neural network, and nonparametric regression models. Evaluation of the impact of new factors in this high dimensional high complexity case will be the subject of future research.

Code Availability

The code to implement the projection-based impact parameter is available using the R package ‘SurvEval,’ which is on CRAN (Devlin and Heller, 2024).

References

- A. R. Brentnall and J. Cuzick. Use of the concordance index for predictors of censored survival data. *Statistical methods in medical research*, 27(8):2359–2373, 2018.
- S. C. Cheng, L. J. Wei, and Z. Ying. Analysis of transformation models with censored data. *Biometrika*, 82(4):835–845, 1995.
- D. M. Dabrowska and K. A. Doksum. Partial likelihood in transformation models with censored data. *Scandinavian Journal of Statistics*, 15(1):1–23, 1988.
- S. Devlin and G. Heller. *SurvEval: Methods for the Evaluation of Survival Models*, 2024. URL <https://CRAN.R-project.org/package=SurvEval>. R package version 1.1.
- S. M. Devlin and G. Heller. Concordance probability as a meaningful contrast across disparate survival times. *Stat Methods Med Res*, 30(3):816–825, Mar 2021.
- J. P. Fine. Comparing nonnested cox models. *Biometrika*, 89(3):635–647, 2002.
- T. A. Gerds. *pec: Prediction Error Curves for Risk Prediction Models in Survival Analysis*, 2023. URL <https://CRAN.R-project.org/package=pec>. R package version 2023.04.12.
- T. A. Gerds, M. W. Kattan, M. Schumacher, and C. Yu. Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring. *Stat Med*, 32(13):2173–2184, Jun 2013.
- P. M. Grambsch and T. M. Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526, 1994.

- M. Gönen and G. Heller. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*, 92:965–970, 2005.
- P. Hall. On projection pursuit regression. *The Annals of Statistics*, 17(2):573–588, 1989.
- F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati. Evaluating the yield of medical tests. *JAMA*, 247(18):2543–2546, May 1982.
- F. E. Harrell, K. L. Lee, and D. B. Mark. Tutorial in biostatistics. multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15:361—387, 1996.
- P. Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396, 1986.
- S. Khan and E. Tamer. Partial rank estimation of of transformation models with general forms of censoring. *Journal of Econometrics*, 136:251–280, 2007.
- M. R. Kosorok, B. L. Lee, and J. P. Fine. Robust inference for univariate proportional hazards frailty regression models. *The Annals of Statistics*, 32(4):1448–1491, 2004.
- M. J. Pencina and R. B. D’Agostino. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med*, 23(13):2109–2123, Jul 2004.
- F. Saad, K. Fizazi, V. Jinga, E. Efstathiou, P. Fong, L. Hart, R. Jones, R. McDermott, M. Wirth, K. Suzuki, D. MacLean, L. Wang, H. Akaza, J. Nelson, H. Scher,

- R. Dreicer, I. Webb, and R. de Wit. Orteronel plus prednisone in patients with chemotherapy naive metastatic castration-resistant prostate cancer (ELM-PC 4): a double-blind, multicentre, phase 3, randomised, placebo-controlled trial. *Lancet Oncology*, 16(3):338–348, March 2015.
- J. S. Simonoff. *Smoothing Methods in Statistics*. Springer Science, New York, 1996.
- X. Song, S. Ma, J. Huang, and X.-H. Zhou. A semiparametric approach for the nonparametric transformation survival model with multiple covariates. *Biostatistics*, 8(2):197–211, 2007.
- H. Uno, T. Cai, M. J. Pencina, R. B. D’Agostino, and L. J. Wei. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*, 30(10):1105–1117, May 2011.
- Y. Zhang and Y. Shao. Concordance measure and discriminatory accuracy in transformation cure models. *Biostatistics*, 19(1):14–26, Jan 2018.

Tables and Figures

Table 1: Proportional hazards estimates for the enhanced model with 631 metastatic prostate cancer patients for the endpoint of overall survival. A total of 248 died during the follow-up period.

	log(RR)	se[log(RR)]	<i>p</i> -value
PSA	0.018	0.006	0.002
Visceral Disease	0.387	0.156	0.013
ECOG Status	-0.408	0.134	0.002
Serum Testosterone	-1.625	0.453	<0.001
CTC _{low}	0.258	0.076	0.001
CTC _{high}	0.106	0.009	<0.001

PSA, testosterone, and CTC were estimated under a square root transformation. CTC_{low} are the transformed values $< \sqrt{10}$, 0 otherwise, while CTC_{high} are values $\geq \sqrt{10}$, 0 otherwise. RR is the relative risk.

Table 2: Proportional hazards estimates for the reduced working model.

	log(RR)	se[log(RR)]	<i>p</i> -value
PSA	0.030	0.004	<0.001
Visceral Disease	0.378	0.154	0.014
ECOG Status	-0.484	0.132	<0.001

PSA was estimated under a square root transformation.

Table 3: Proportional hazards simulation results for the impact parameter equal to 0.025. The average bias and relative efficiency of the projection-based CPE with respect to the weighted c-index were averaged over 2,000 simulation iterations. Standard errors were calculated based on 50 bootstrap samples within each iteration and compared to the simulation standard error. Coverage was based on 95% confidence intervals using the bootstrap standard error estimate.

	N	Enhanced	Projection	Estimate	Censoring	Bias	Relative Efficiency CPE w.r.t. wCI	Estimated to Simulation SE	Coverage
Projection									
	300	0.70	0.675	PL/CPE	0%	0.001	1.000	0.991	0.949
	300	0.70	0.675	PL/wCI	0%	<0.001	1.225	1.010	0.951
	300	0.70	0.675	PR/wCI	0%	<0.001	1.225	1.010	0.951
	300	0.70	0.675	PL/CPE	25%	0.001	1.000	0.999	0.953
	300	0.70	0.675	PL/wCI	25%	0.000	1.164	1.013	0.954
	300	0.70	0.675	PR/wCI	25%	0.001	1.165	1.013	0.952
	300	0.70	0.675	PL/CPE	50%	0.001	1.000	0.972	0.939
	300	0.70	0.675	PL/wCI	50%	0.001	1.094	1.003	0.953
	300	0.70	0.675	PR/wCI	50%	0.001	1.096	1.007	0.951
Impact									
	300	0.70	0.675	PL/CPE	0%	0.001	1.000	1.006	0.948
	300	0.70	0.675	PL/wCI	0%	0.001	1.324	1.028	0.957
	300	0.70	0.675	PR/wCI	0%	0.002	1.344	1.034	0.953
	300	0.70	0.675	PL/CPE	25%	0.002	1.000	1.014	0.949
	300	0.70	0.675	PL/wCI	25%	0.002	1.212	1.040	0.958
	300	0.70	0.675	PR/wCI	25%	0.002	1.227	1.045	0.954
	300	0.70	0.675	PL/CPE	50%	0.003	1.000	1.019	0.946
	300	0.70	0.675	PL/wCI	50%	0.003	1.108	1.072	0.959
	300	0.70	0.675	PR/wCI	50%	0.003	1.116	1.081	0.961

Table 4: Proportional hazards simulation results for the impact parameter equal to 0.05. The average bias and relative efficiency of the projection-based CPE with respect to the weighted c-index were averaged over 2,000 simulation iterations. Standard errors were calculated based on 50 bootstrap samples within each iteration and compared to the simulation standard error. Coverage was based on 95% confidence intervals using the bootstrap standard error estimate.

	N	Enhanced	Projection	Estimate	Censoring	Bias	Relative Efficiency CPE w.r.t. wCI	Estimated to Simulation SE	Coverage
Projection									
	300	0.70	0.65	PL/CPE	0%	0.002	1.000	1.040	0.959
	300	0.70	0.65	PL/wCI	0%	0.001	1.238	1.018	0.951
	300	0.70	0.65	PR/wCI	0%	0.001	1.238	1.018	0.954
	300	0.70	0.65	PL/CPE	25%	0.002	1.000	1.003	0.950
	300	0.70	0.65	PL/wCI	25%	0.001	1.176	1.012	0.950
	300	0.70	0.65	PR/wCI	25%	0.001	1.177	1.014	0.949
	300	0.70	0.65	PL/CPE	50%	0.002	1.000	0.980	0.941
	300	0.70	0.65	PL/wCI	50%	0.002	1.101	1.008	0.949
	300	0.70	0.65	PR/wCI	50%	0.002	1.104	1.013	0.952
Impact									
	300	0.70	0.65	PL/CPE	0%	0.001	1.000	1.020	0.956
	300	0.70	0.65	PL/wCI	0%	0.001	1.304	1.019	0.956
	300	0.70	0.65	PR/wCI	0%	0.002	1.314	1.021	0.957
	300	0.70	0.65	PL/CPE	25%	0.001	1.000	1.012	0.949
	300	0.70	0.65	PL/wCI	25%	0.001	1.230	1.015	0.954
	300	0.70	0.65	PR/wCI	25%	0.002	1.240	1.016	0.955
	300	0.70	0.65	PL/CPE	50%	0.002	1.000	1.012	0.957
	300	0.70	0.65	PL/wCI	50%	0.002	1.121	1.041	0.959
	300	0.70	0.65	PR/wCI	50%	0.002	1.126	1.047	0.959

Table 5: Proportional hazards simulation results for the impact parameter equal to 0.10. The average bias and relative efficiency with respect to the projection-based CPE were averaged over 2,000 simulation iterations. Standard errors were calculated based on 50 bootstrap samples within each iteration and compared to the simulation standard error. Coverage was based on 95% confidence intervals using the bootstrap standard error estimate.

	N	Enhanced	Projection	Estimate	Censoring	Bias	Relative Efficiency CPE w.r.t. wCI	Estimated to Simulation SE	Coverage
Projection									
	300	0.70	0.60	PL/CPE	0%	0.001	1.000	1.006	0.954
	300	0.70	0.60	PL/wCI	0%	0.001	1.241	1.034	0.957
	300	0.70	0.60	PR/wCI	0%	0.001	1.240	1.038	0.959
	300	0.70	0.60	PL/CPE	25%	0.001	1.000	1.009	0.959
	300	0.70	0.60	PL/wCI	25%	0.001	1.167	1.033	0.958
	300	0.70	0.60	PR/wCI	25%	0.002	1.169	1.035	0.957
	300	0.70	0.60	PL/CPE	50%	0.002	1.000	0.996	0.948
	300	0.70	0.60	PL/wCI	50%	0.002	1.104	1.024	0.958
	300	0.70	0.60	PR/wCI	50%	0.002	1.108	1.032	0.960
Impact									
	300	0.70	0.60	PL/CPE	0%	0.001	1.000	0.987	0.945
	300	0.70	0.60	PL/wCI	0%	0.001	1.248	1.028	0.952
	300	0.70	0.60	PR/wCI	0%	0.002	1.251	1.026	0.952
	300	0.70	0.60	PL/CPE	25%	0.001	1.000	1.000	0.953
	300	0.70	0.60	PL/wCI	25%	0.001	1.196	1.016	0.951
	300	0.70	0.60	PR/wCI	25%	0.002	1.201	1.014	0.950
	300	0.70	0.60	PL/CPE	50%	0.002	1.000	1.010	0.953
	300	0.70	0.60	PL/wCI	50%	0.001	1.109	1.044	0.955
	300	0.70	0.60	PR/wCI	50%	0.002	1.110	1.051	0.958

Table 6: Non-proportional hazards simulation results for the impact parameter equal to 0.025. The average bias and relative efficiency with respect to the projection-based CPE were averaged over 2,000 simulation iterations. Standard errors were calculated based on 50 bootstrap samples within each iteration and compared to the simulation standard error. Coverage was based on 95% confidence intervals using the bootstrap standard error estimate.

	N	Enhanced	Projection	Estimate	Censoring	Bias	Relative Efficiency CPE w.r.t. wCI	Estimated to Simulation SE	Coverage
Projection									
	300	0.70	0.675	PL/CPE	0%	-0.107	1.000	0.737	0.006
	300	0.70	0.675	PL/wCI	0%	-0.008	0.333	0.802	0.958
	300	0.70	0.675	PR/wCI	0%	0.001	0.192	1.001	0.952
	300	0.70	0.675	PL/CPE	25%	-0.053	1.000	0.959	0.270
	300	0.70	0.675	PL/wCI	25%	0.001	0.377	1.009	0.942
	300	0.70	0.675	PR/wCI	25%	0.002	0.376	0.999	0.940
	300	0.70	0.675	PL/CPE	50%	-0.016	1.000	0.965	0.875
	300	0.70	0.675	PL/wCI	50%	0.002	0.794	1.004	0.946
	300	0.70	0.675	PR/wCI	50%	0.002	0.795	1.003	0.945
Impact									
	300	0.70	0.675	PL/CPE	0%	-0.013	1.000	1.010	0.607
	300	0.70	0.675	PL/wCI	0%	-0.006	1.444	0.869	0.936
	300	0.70	0.675	PR/wCI	0%	0.002	0.741	1.077	0.959
	300	0.70	0.675	PL/CPE	25%	-0.007	1.000	1.012	0.805
	300	0.70	0.675	PL/wCI	25%	<0.001	0.984	1.091	0.945
	300	0.70	0.675	PR/wCI	25%	0.002	0.981	1.066	0.949
	300	0.70	0.675	PL/CPE	50%	-0.001	1.000	1.022	0.909
	300	0.70	0.675	PL/wCI	50%	0.001	1.033	1.065	0.946
	300	0.70	0.675	PR/wCI	50%	0.003	1.051	1.059	0.948

Table 7: Non-proportional hazards simulation results for the impact parameter equal to 0.05. The average bias and relative efficiency with respect to the projection-based CPE were averaged over 2,000 simulation iterations. Standard errors were calculated based on 50 bootstrap samples within each iteration and compared to the simulation standard error. Coverage was based on 95% confidence intervals using the bootstrap standard error estimate.

	N	Enhanced	Projection	Estimate	Censoring	Bias	Relative Efficiency CPE w.r.t. wCI	Estimated to Simulation SE	Coverage
Projection									
	300	0.70	0.65	PL/CPE	0%	-0.092	1.000	0.763	0.021
	300	0.70	0.65	PL/wCI	0%	-0.008	0.396	0.799	0.952
	300	0.70	0.65	PR/wCI	0%	0.001	0.227	1.007	0.952
	300	0.70	0.65	PL/CPE	25%	-0.044	1.000	0.975	0.387
	300	0.70	0.65	PL/wCI	25%	0.001	0.450	1.020	0.954
	300	0.70	0.65	PR/wCI	25%	0.001	0.449	1.006	0.951
	300	0.70	0.65	PL/CPE	50%	-0.012	1.000	0.987	0.913
	300	0.70	0.65	PL/wCI	50%	0.001	0.875	1.015	0.953
	300	0.70	0.65	PR/wCI	50%	0.002	0.877	1.013	0.954
Impact									
	300	0.70	0.65	PL/CPE	0%	-0.031	1.000	0.934	0.254
	300	0.70	0.65	PL/wCI	0%	-0.006	0.809	0.868	0.938
	300	0.70	0.65	PR/wCI	0%	0.002	0.470	1.058	0.959
	300	0.70	0.65	PL/CPE	25%	-0.016	1.000	1.002	0.772
	300	0.70	0.65	PL/wCI	25%	<0.001	0.746	1.079	0.963
	300	0.70	0.65	PR/wCI	25%	0.002	0.746	1.059	0.959
	300	0.70	0.65	PL/CPE	50%	-0.005	1.000	1.029	0.952
	300	0.70	0.65	PL/wCI	50%	0.001	0.984	1.064	0.964
	300	0.70	0.65	PR/wCI	50%	0.003	0.992	1.060	0.962

Table 8: Non-proportional hazards simulation results for the impact parameter equal to 0.10. The average bias and relative efficiency with respect to the projection-based CPE were averaged over 2,000 simulation iterations. Standard errors were calculated based on 50 bootstrap samples within each iteration and compared to the simulation standard error. Coverage was based on 95% confidence intervals using the bootstrap standard error estimate.

	N	Enhanced	Projection	Estimate	Censoring	Bias	Relative Efficiency CPE w.r.t. wCI	Estimated to Simulation SE	Coverage
Projection									
	300	0.70	0.65	PL/CPE	0%	-0.058	1.000	0.861	0.146
	300	0.70	0.65	PL/wCI	0%	-0.010	0.627	0.842	0.932
	300	0.70	0.65	PR/wCI	0%	0.002	0.365	1.018	0.954
	300	0.70	0.60	PL/CPE	25%	-0.027	1.000	0.986	0.719
	300	0.70	0.60	PL/wCI	25%	0.001	0.683	1.050	0.959
	300	0.70	0.60	PR/wCI	25%	0.002	0.678	1.019	0.954
	300	0.70	0.60	PL/CPE	50%	-0.006	1.000	1.012	0.949
	300	0.70	0.60	PL/wCI	50%	0.002	0.986	1.039	0.962
	300	0.70	0.60	PR/wCI	50%	0.002	0.992	1.030	0.958
Impact									
	300	0.70	0.60	PL/CPE	0%	-0.066	1.000	0.855	0.071
	300	0.70	0.60	PL/wCI	0%	-0.004	0.509	0.889	0.936
	300	0.70	0.60	PR/wCI	0%	0.002	0.320	1.047	0.959
	300	0.70	0.60	PL/CPE	25%	-0.034	1.000	0.977	0.569
	300	0.70	0.60	PL/wCI	25%	<0.001	0.568	1.073	0.965
	300	0.70	0.60	PR/wCI	25%	0.002	0.562	1.045	0.959
	300	0.70	0.60	PL/CPE	50%	-0.012	1.000	1.016	0.928
	300	0.70	0.60	PL/wCI	50%	0.001	0.901	1.057	0.962
	300	0.70	0.60	PR/wCI	50%	0.003	0.901	1.052	0.957

Supplementary Material

Measuring the Impact of New Risk Factors Within Survival

Models

S1 Assessment of the proportional hazards assumption
in data example of 631 metastatic prostate cancer
patients.

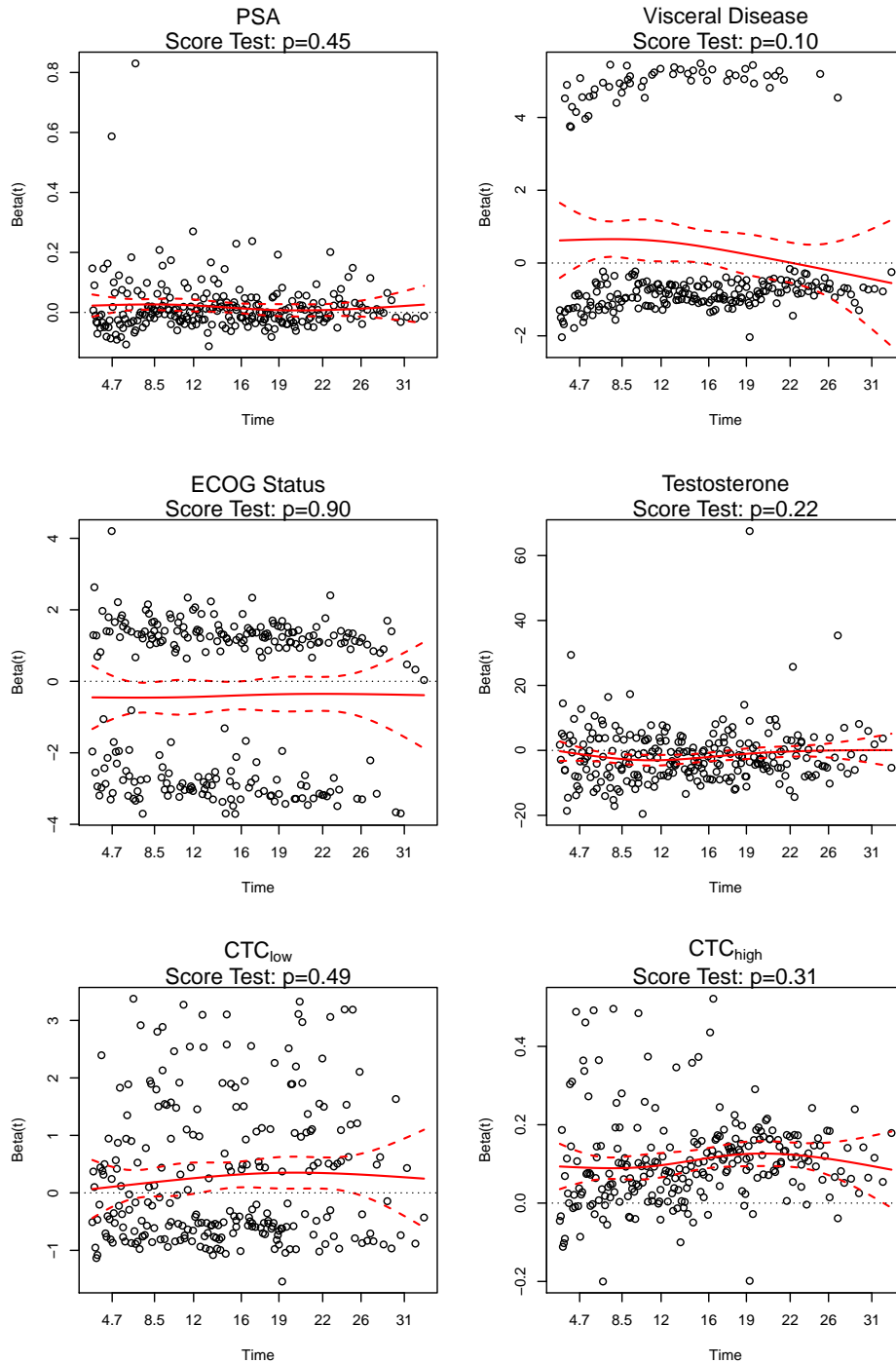


Figure S1: The smoothed relationship between the scaled Schoenfeld residuals and time are plotted for each risk factor in the enhanced model. The p-value provided at the top of each plot is generated from a null test statistic of a constant slope.

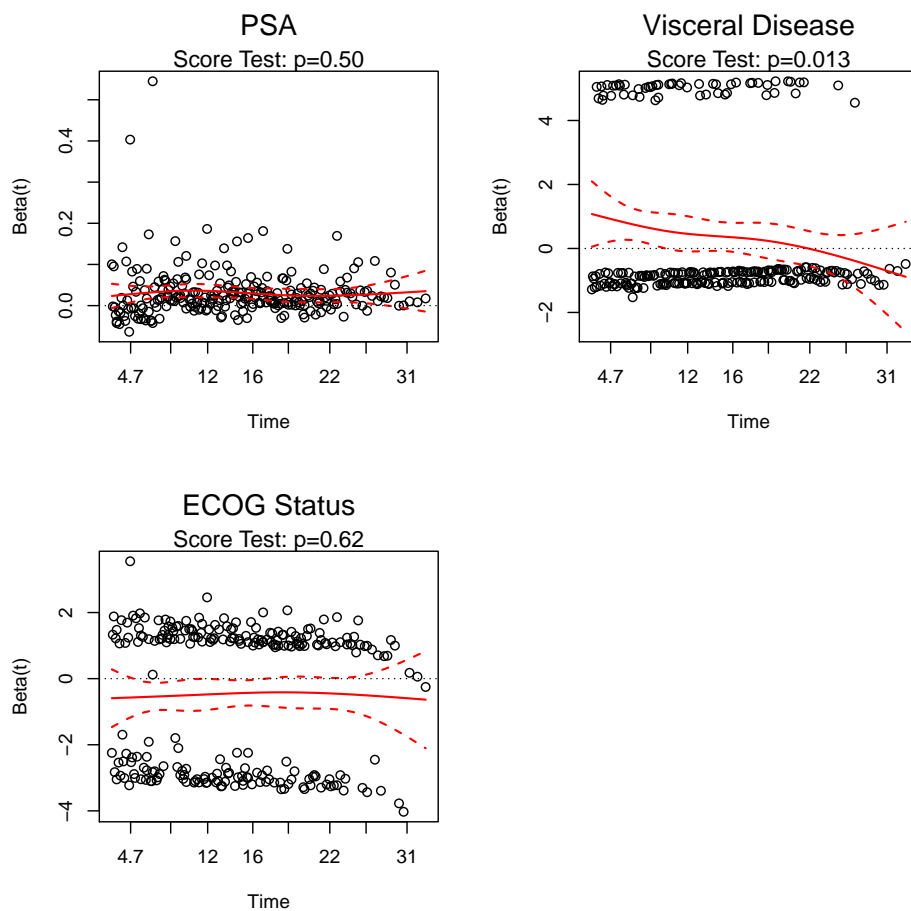


Figure S2: The Schoenfeld residuals are plotted over time for each risk factor in the reduced model. The p-value provided at the top of each plot is generated from a null test statistic of a constant slope.

S2 Derivation of Theorems 1 and 2

The following notation and regularity conditions are used throughout the appendix.

(C1) The difference in risk indices are denoted by

$$\boldsymbol{\beta}^T \mathbf{x}_{ij} = \boldsymbol{\beta}^T \mathbf{x}_i - \boldsymbol{\beta}^T \mathbf{x}_j, \quad \boldsymbol{\gamma}^T \mathbf{z}_{ij} = \boldsymbol{\gamma}^T \mathbf{z}_i - \boldsymbol{\gamma}^T \mathbf{z}_j$$

and the marginal precedence probability is represented as

$$\pi(\tau) = \Pr(T_1 < T_2, T_1 < \tau).$$

(C2) $(T_i, \mathbf{X}_i, \mathbf{Z}_i), i = 1, \dots, n$ are an independent random sample from $(T, \mathbf{X}, \mathbf{Z})$.

(C3) The random vector (\mathbf{X}, \mathbf{Z}) and the parameter vector $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ each lie in a $p+q$ dimensional bounded rectangle, where $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ are the true values of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$.

(C4) The enhanced model precedence probability $\theta_{abcd}(\boldsymbol{\beta}, \boldsymbol{\gamma}; \tau) =$

$$\Pr(T_1 < T_2, T_1 < \tau | \boldsymbol{\beta}^T \mathbf{X}_1 = \boldsymbol{\beta}^T \mathbf{x}_a, \boldsymbol{\beta}^T \mathbf{X}_2 = \boldsymbol{\beta}^T \mathbf{x}_b, \boldsymbol{\gamma}^T \mathbf{Z}_1 = \boldsymbol{\gamma}^T \mathbf{z}_c, \boldsymbol{\gamma}^T \mathbf{Z}_2 = \boldsymbol{\gamma}^T \mathbf{z}_d)$$

has bounded first partial derivatives in a compact neighborhood of $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$.

(C5) As $n \rightarrow \infty$, the centered estimates

$$n^{1/2} \left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0), (\hat{\theta}_{abcd}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0; \tau) - \theta_{abcd}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0; \tau)), (\hat{\pi}(\tau) - \pi_0(\tau)) \right]$$

converge to a multivariate normal random vector with mean zero.

(C6) $\phi(u/h)$ and $\Phi(u/g)$ are a normal kernel density function with bandwidth h and a normal local distribution function with bandwidth g , respectively.

As $n \rightarrow \infty$, the bandwidths satisfy, $h, g \rightarrow 0$, $nh^2 \rightarrow \infty$, $nh^4 \rightarrow 0$,

and $ng^4 \rightarrow 0$.

Theorem 1

$$\text{Let } \hat{\theta}_{12}^{[P]}(\hat{\boldsymbol{\beta}}; \tau) = \frac{n^{-2} \sum_{k,l} \hat{\theta}_{12kl}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}; \tau) \phi_h(\hat{\boldsymbol{\beta}}^T \mathbf{x}_1, \hat{\boldsymbol{\beta}}^T \mathbf{x}_k) \phi_h(\hat{\boldsymbol{\beta}}^T \mathbf{x}_2, \hat{\boldsymbol{\beta}}^T \mathbf{x}_l)}{n^{-2} \sum_{k,l} \phi_h(\hat{\boldsymbol{\beta}}^T \mathbf{x}_1, \hat{\boldsymbol{\beta}}^T \mathbf{x}_k) \phi_h(\hat{\boldsymbol{\beta}}^T \mathbf{x}_2, \hat{\boldsymbol{\beta}}^T \mathbf{x}_l)}$$

$$\text{Then } \hat{\theta}_{12}^{[P]}(\hat{\boldsymbol{\beta}}; \tau) = \Pr(T_1 < T_2, T_1 < \tau | \boldsymbol{\beta}_0^T \mathbf{x}_1, \boldsymbol{\beta}_0^T \mathbf{x}_2) + o_p(1).$$

Proof:

The coefficient estimate $\hat{\boldsymbol{\beta}}$ within the kernel $\phi(\cdot)$ may be treated as a constant since its \sqrt{n} convergence rate is faster than the nonparametric kernel estimate rate.

The denominator of the Nadarya-Watson kernel estimate $\hat{\theta}_{12}^{[P]}$, may be represented asymptotically as

$$n^{-2} \sum_{k,l} \phi_h(\hat{\boldsymbol{\beta}}^T \mathbf{x}_1, \hat{\boldsymbol{\beta}}^T \mathbf{x}_k) \phi_h(\hat{\boldsymbol{\beta}}^T \mathbf{x}_2, \hat{\boldsymbol{\beta}}^T \mathbf{x}_l) = f_{\beta_0^T X_1}(\boldsymbol{\beta}_0^T \mathbf{x}_1) f_{\beta_0^T X_2}(\boldsymbol{\beta}_0^T \mathbf{x}_2) + o_p(1).$$

For the numerator, a first order Taylor series expansion of the model precedence probability around $\left[\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0, \hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}_0, \hat{\theta}_{abcd}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0; \tau) = \theta_{abcd}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0; \tau) \right]$ results in

$$\begin{aligned} & n^{-2} \sum_{k,l} \hat{\theta}_{12kl}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}; \tau) \phi_h(\hat{\boldsymbol{\beta}}^T \mathbf{x}_1, \hat{\boldsymbol{\beta}}^T \mathbf{x}_k) \phi_h(\hat{\boldsymbol{\beta}}^T \mathbf{x}_2, \hat{\boldsymbol{\beta}}^T \mathbf{x}_l) \\ &= \Pr(T_1 < T_2, T_1 < \tau | \boldsymbol{\beta}_0^T \mathbf{x}_1, \boldsymbol{\beta}_0^T \mathbf{x}_2) f_{\beta_0^T X_1}(\boldsymbol{\beta}_0^T \mathbf{x}_1) f_{\beta_0^T X_2}(\boldsymbol{\beta}_0^T \mathbf{x}_2) + o_p(1). \end{aligned}$$

Therefore, $\hat{\theta}_{12}^{[P]}(\hat{\boldsymbol{\beta}}; \tau) = \Pr(T_1 < T_2, T_1 < \tau | \boldsymbol{\beta}_0^T \mathbf{x}_1, \boldsymbol{\beta}_0^T \mathbf{x}_2) + o_p(1)$.

Theorem 2:

The following notation, used throughout the proof of Theorem 2, is collected here.

Consider the concordance probability estimate (CPE) from the extended model and the CPE for \mathbf{X} alone obtained by projection

$$K_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}; \tau), \hat{\pi}(\tau)) = [n^2 \hat{\pi}(\tau)]^{-1} \sum_{i,j} I(\hat{\boldsymbol{\beta}}^T \mathbf{x}_{ij} + \hat{\boldsymbol{\gamma}}^T \mathbf{z}_{ij} > 0) \hat{\theta}_{ijij}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}; \tau)$$

$$K_n^{[P]}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}^{[P]}(\hat{\boldsymbol{\beta}}; \tau), \hat{\pi}(\tau)) = [n^2 \hat{\pi}(\tau)]^{-1} \sum_{i,j} I(\hat{\boldsymbol{\beta}}^T \mathbf{x}_{ij} > 0) \hat{\theta}_{ij}^{[P]}(\hat{\boldsymbol{\beta}}; \tau),$$

where to make the arguments in the CPE statistics more concise, going forward we write

$$\hat{\theta} \equiv \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}; \tau) \quad \hat{\theta}_0 \equiv \hat{\boldsymbol{\theta}}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0; \tau) \quad \theta_0 \equiv \boldsymbol{\theta}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0; \tau) \quad \hat{\pi}(\tau) \equiv \hat{\pi} \quad \pi_0(\tau) \equiv \pi_0.$$

The impact parameter is denoted by

$$\xi(\tau) = \kappa(\tau) - \kappa^{[P]}(\tau),$$

where as described by equations (5) and (9) in the text,

$$K_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}, \hat{\pi}; \tau) \xrightarrow{P} \kappa(\tau)$$

$$K_n^{[P]}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}^{[P]}, \hat{\pi}; \tau) \xrightarrow{P} \kappa^{[P]}(\tau).$$

Then Theorem 2 is stated as

$$n^{1/2} \left[K_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}, \hat{\pi}; \tau) - K_n^{[P]}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}^{[P]}, \hat{\pi}; \tau) - \xi(\tau) \right] \xrightarrow{D} N(0, V) \quad (\text{S.1})$$

Proof

To derive this result, smooth versions of K_n and $K_n^{[P]}$ are employed

$$\tilde{K}_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\pi}; \tau) = [n^2 \hat{\pi}(\tau)]^{-1} \sum_{i,j} \Phi \left(\frac{\hat{\beta}^T \mathbf{x}_{ij} + \hat{\gamma}^T \mathbf{z}_{ij}}{g} \right) \hat{\theta}_{ijij}(\hat{\beta}, \hat{\gamma}; \tau)$$

$$\tilde{K}_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \hat{\pi}; \tau) = [n^2 \hat{\pi}(\tau)]^{-1} \sum_{i,j} \Phi \left(\frac{\hat{\beta}^T \mathbf{x}_{ij}}{g} \right) \hat{\theta}_{ij}^{[P]}(\hat{\beta}; \tau)$$

The asymptotic normal distribution (S.1) is demonstrated via the decomposition

$$\begin{aligned} & n^{1/2} \left\{ \left[K_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\pi}; \tau) - \tilde{K}_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\pi}; \tau) \right] - \left[K_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \hat{\pi}; \tau) - \tilde{K}_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \hat{\pi}; \tau) \right] \right\} + \\ & n^{1/2} \left\{ \left[\tilde{K}_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\pi}; \tau) - \tilde{K}_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \pi_0; \tau) \right] - \left[\tilde{K}_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \hat{\pi}; \tau) - \tilde{K}_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \pi_0; \tau) \right] \right\} + \\ & n^{1/2} \left\{ \left[\tilde{K}_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \pi_0; \tau) - \tilde{K}_n(\beta_0, \gamma_0, \hat{\theta}_0, \pi_0; \tau) \right] - \left[\tilde{K}_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \pi_0; \tau) - \tilde{K}_n^{[P]}(\beta_0, \hat{\theta}_0^{[P]}, \pi_0; \tau) \right] \right\} + \\ & n^{1/2} \left\{ \left[\tilde{K}_n(\beta_0, \gamma_0, \hat{\theta}_0, \pi_0; \tau) - \tilde{K}_n(\beta_0, \gamma_0, \theta_0, \pi_0; \tau) \right] - \left[\tilde{K}_n^{[P]}(\beta_0, \hat{\theta}_0^{[P]}, \pi_0; \tau) - \tilde{K}_n^{[P]}(\beta_0, \theta_0^{[P]}, \pi_0; \tau) \right] \right\} + \\ & n^{1/2} \left\{ \left[\tilde{K}_n(\beta_0, \gamma_0, \theta_0, \pi_0; \tau) - \kappa(\tau) \right] - \left[\tilde{K}_n^{[P]}(\beta_0, \theta_0, \pi_0; \tau) - \kappa^{[P]}(\tau) \right] \right\}. \end{aligned}$$

The asymptotic results are developed below.

- From the first line of the decomposition,

$$\begin{aligned} & n^{1/2} \left\{ \left[K_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\pi}; \tau) - \tilde{K}_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\pi}; \tau) \right] \right\} = \\ & n^{-3/2} [\hat{\pi}(\tau)]^{-1} \sum_{i,j} \left\{ I(\hat{\beta}^T \mathbf{x}_{ij} + \hat{\gamma}^T \mathbf{z}_{ij} > 0) - \Phi \left(\frac{\hat{\beta}^T \mathbf{x}_{ij} + \hat{\gamma}^T \mathbf{z}_{ij}}{g} \right) \right\} \hat{\theta}_{ijij}(\hat{\beta}, \hat{\gamma}; \tau). \end{aligned}$$

Since $0 < \hat{\pi}(\tau) < 1$ and $\left| \hat{\theta}_{ijij}(\hat{\beta}, \hat{\gamma}; \tau) \right| < 1$,

$$< Mn^{-3/2} \sup_{\beta, \gamma} \left| \sum_{i,j} \left\{ I(\beta^T \mathbf{x}_{ij} + \gamma^T \mathbf{z}_{ij} > 0) - \Phi \left(\frac{\beta^T \mathbf{x}_{ij} + \gamma^T \mathbf{z}_{ij}}{g} \right) \right\} \right|$$

The right hand side of this inequality may be rewritten as

$$Mn^{1/2} \sup_{\beta, \gamma} \left| \int_u \left\{ I(u > 0) - \Phi \left(\frac{u}{g} \right) \right\} d\hat{F}_{n \times n}(u) \right|$$

where $\hat{F}_{n \times n}(u)$ is the empirical cumulative distribution function with jumps at each of the n^2 elements of u_{ij} . This expression has the same form as the expression in Heller (2007, Lemma A.1), where it is shown

$$n^{1/2} \left\{ \left[K_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\pi}; \tau) - \tilde{K}_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\pi}; \tau) \right] \right\} \xrightarrow{p} 0$$

uniformly in (β, γ) .

The same argument demonstrates that

$$n^{1/2} \left\{ \left[K_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \hat{\pi}; \tau) - \tilde{K}_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \hat{\pi}; \tau) \right] \right\} \xrightarrow{p} 0$$

uniformly in (β, γ) .

- For the second line of the decomposition

$$n^{1/2} \left\{ \left[\tilde{K}_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\pi}; \tau) - \tilde{K}_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \pi_0; \tau) \right] \right\} =$$

$$\left\{ n^{1/2} \frac{[\pi_0(\tau) - \hat{\pi}(\tau)]}{\pi_0(\tau) \hat{\pi}(\tau)} \right\} \left\{ n^{-2} \sum_{i,j} \Phi \left(\frac{\hat{\beta}^T \mathbf{x}_{ij} + \hat{\gamma}^T \mathbf{z}_{ij}}{g} \right) \hat{\theta}_{ijij}(\hat{\beta}, \hat{\gamma}; \tau) \right\}$$

The first term in curly brackets, from Kaplan-Meier estimation theory, converges in distribution to a mean zero normal random variable. The second term in curly brackets converges in probability to a constant. It follows from Slutsky's Theorem that

$$n^{1/2} \left\{ \left[\tilde{K}_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \hat{\pi}; \tau) - \tilde{K}_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \pi_0; \tau) \right] \right\} \xrightarrow{D} N(0, V_2).$$

A similar argument gives

$$n^{1/2} \left\{ \left[K_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \hat{\pi}; \tau) - \tilde{K}_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \pi_0; \tau) \right] \right\} \xrightarrow{D} N(0, V_{2P}).$$

- In the third and fourth lines of the decomposition, using the asymptotic normality

in (C5), a first order Taylor series around $[\hat{\beta}, \hat{\gamma}, \hat{\theta}(\hat{\beta}, \hat{\gamma})] = [\beta_0, \gamma_0, \theta_0]$, results in

$$\begin{aligned} n^{1/2} \left\{ \tilde{K}_n(\hat{\beta}, \hat{\gamma}, \hat{\theta}, \pi_0; \tau) - \tilde{K}_n(\beta_0, \gamma_0, \hat{\theta}_0, \pi_0; \tau) \right\} &\xrightarrow{D} N(0, V_3) \\ n^{1/2} \left\{ \tilde{K}_n^{[P]}(\hat{\beta}, \hat{\theta}^{[P]}, \pi_0; \tau) - \tilde{K}_n^{[P]}(\beta_0, \hat{\theta}_0^{[P]}, \pi_0; \tau) \right\} &\xrightarrow{D} N(0, V_{3P}) \\ n^{1/2} \left\{ \tilde{K}_n(\beta_0, \gamma_0, \hat{\theta}_0, \pi_0; \tau) - \tilde{K}_n(\beta_0, \gamma_0, \theta_0, \pi_0; \tau) \right\} &\xrightarrow{D} N(0, V_4) \\ n^{1/2} \left\{ \tilde{K}_n^{[P]}(\beta_0, \hat{\theta}_0^{[P]}, \pi_0; \tau) - \tilde{K}_n^{[P]}(\beta_0, \gamma_0, \theta_0^{[P]}, \pi_0; \tau) \right\} &\xrightarrow{D} N(0, V_{4P}) \end{aligned}$$

- From the fifth line of the decomposition,

$$\begin{aligned} n^{1/2} \left[\tilde{K}_n(\beta_0, \gamma_0, \theta_0, \pi_0; \tau) - \kappa(\tau) \right] \\ = n^{-3/2} \sum_{i,j} \left\{ [\pi_0(\tau)]^{-1} \left[\Phi \left(\frac{\beta^T \mathbf{x}_{ij}}{g} \right) \theta_{ijij}(\beta_0, \gamma_0; \tau) \right] - \kappa(\tau) \right\} \end{aligned}$$

is a degree 2 U-statistic, and therefore

$$n^{1/2} \left[\tilde{K}_n(\beta_0, \gamma_0, \theta_0, \pi_0; \tau) - \kappa(\tau) \right] \xrightarrow{D} N(0, V_5).$$

For the projected difference

$$\begin{aligned} n^{1/2} \left[\tilde{K}_n^{[P]}(\beta_0, \theta_0^{[P]}, \pi_0; \tau) - \kappa^{[P]}(\tau) \right] = \\ n^{-3/2} \sum_{i,j} \left[\Phi \left(\frac{\beta_0^T \mathbf{x}_{ij}}{g} \right) \left\{ \frac{\sum_{k,l} \theta_{ijkl}(\beta_0, \gamma_0; \tau) \phi_h(\beta_0^T \mathbf{x}_i, \beta_0^T \mathbf{x}_k) \phi_h(\beta_0^T \mathbf{x}_j, \beta_0^T \mathbf{x}_l)}{\pi_0(\tau) \sum_{k,l} \phi_h(\beta_0^T \mathbf{x}_i, \beta_0^T \mathbf{x}_k) \phi_h(\beta_0^T \mathbf{x}_j, \beta_0^T \mathbf{x}_l)} \right\} - \kappa^{[P]}(\tau) \right]. \end{aligned}$$

$$\text{Let } a_{1ij}(\beta_0; \tau) = n^{-2} \sum_{k,l} \theta_{ijkl}(\beta_0, \gamma_0; \tau) \phi_h(\beta_0^T \mathbf{x}_i, \beta_0^T \mathbf{x}_k) \phi_h(\beta_0^T \mathbf{x}_j, \beta_0^T \mathbf{x}_l)$$

$$a_{0ij}(\beta_0; \tau) = n^{-2} \sum_{k,l} \phi_h(\beta_0^T \mathbf{x}_i, \beta_0^T \mathbf{x}_k) \phi_h(\beta_0^T \mathbf{x}_j, \beta_0^T \mathbf{x}_l)$$

$$\alpha_{mij}(\beta_0; \tau) = \lim_{n \rightarrow \infty} a_{mij}(\beta_0; \tau) \quad m = 0, 1$$

Then

$$n^{1/2} \left[\tilde{K}_n^{[P]}(\boldsymbol{\beta}_0, \theta_0^{[P]}, \pi_0; \tau) - \kappa^{[P]}(\tau) \right] = n^{-3/2} \sum_{i,j} \left[\frac{\Phi \left(\frac{\boldsymbol{\beta}_0^T \mathbf{x}_{ij}}{g} \right) a_{1ij}(\boldsymbol{\beta}_0; \tau)}{\pi_0(\tau) a_{0ij}(\boldsymbol{\beta}_0; \tau)} - \kappa^{[P]}(\tau) \right]$$

Taylor expanding the ratio $a_{1ij}(\boldsymbol{\beta}_0; \tau)/a_{0ij}(\boldsymbol{\beta}_0; \tau)$ around $\alpha_{1ij}(\boldsymbol{\beta}_0; \tau)/\alpha_{0ij}(\boldsymbol{\beta}_0; \tau)$,

$$n^{1/2} \left[\tilde{K}_n^{[P]}(\boldsymbol{\beta}_0, \theta_0^{[P]}, \pi; \tau) - \kappa^{[P]}(\tau) \right] = n^{-7/2} \times \sum_{i,j,k,l} \left(\Phi \left(\frac{\boldsymbol{\beta}_0^T \mathbf{x}_{ij}}{g} \right) \left\{ \frac{\alpha_{1ij}(\boldsymbol{\beta}_0; \tau)}{\pi_0(\tau) \alpha_{0ij}(\boldsymbol{\beta}_0; \tau)} + \frac{\phi_h(\boldsymbol{\beta}_0^T \mathbf{x}_i, \boldsymbol{\beta}_0^T \mathbf{x}_k) \phi_h(\boldsymbol{\beta}_0^T \mathbf{x}_j, \boldsymbol{\beta}_0^T \mathbf{x}_l)}{\pi_0(\tau) \alpha_{0ij}(\boldsymbol{\beta}_0; \tau)} \left[\theta_{ijkl}(\boldsymbol{\beta}_0, \gamma_0; \tau) - \frac{\alpha_{1ij}(\boldsymbol{\beta}_0; \tau)}{\alpha_{0ij}(\boldsymbol{\beta}_0; \tau)} \right] \right\} - \kappa^{[P]}(\tau) \right)$$

which is a degree 4 U-statistic and therefore

$$n^{1/2} [\tilde{K}_n^{[P]}(\boldsymbol{\beta}_0, \theta_0^{[P]}, \pi; \tau) - \kappa^{[P]}(\tau)] \xrightarrow{D} N(0, V_{5P}).$$

Therefore, combining the terms of the decomposition,

$$n^{1/2} \left[K_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\theta}, \hat{\pi}; \tau) - K_n^{[P]}(\hat{\boldsymbol{\beta}}, \hat{\theta}^{[P]}, \hat{\pi}; \tau) - \xi(\tau) \right] \xrightarrow{D} N(0, V).$$

S3 Assessment of the proportional hazards assumption in the simulation scenarios

Table S1: The proportion of times the proportional hazards test proposed by Grambsch and Therneau (1994) for the enhanced model was rejected at the 0.05 level across all simulation scenarios. The proportion is calculated over the 2,000 simulation iterations in each scenario.

Simulation Setting	Enhanced	Projection	Censoring	PH Test
				Proportion $P < 0.05$
Proportional Hazards	0.70	0.675	0%	0.050
Proportional Hazards	0.70	0.675	25%	0.056
Proportional Hazards	0.70	0.675	50%	0.056
Proportional Hazards	0.70	0.65	0%	0.065
Proportional Hazards	0.70	0.65	25%	0.055
Proportional Hazards	0.70	0.65	50%	0.053
Proportional Hazards	0.70	0.60	0%	0.067
Proportional Hazards	0.70	0.60	25%	0.053
Proportional Hazards	0.70	0.60	50%	0.057
Non-proportional Hazards	0.70	0.675	0%	1.00
Non-proportional Hazards	0.70	0.675	25%	0.998
Non-proportional Hazards	0.70	0.675	50%	0.922
Non-proportional Hazards	0.70	0.65	0%	1.00
Non-proportional Hazards	0.70	0.65	25%	0.998
Non-proportional Hazards	0.70	0.65	50%	0.933
Non-proportional Hazards	0.70	0.60	0%	1.00
Non-proportional Hazards	0.70	0.60	25%	0.999
Non-proportional Hazards	0.70	0.60	50%	0.927