

## Article

# Analysis of Receiver Operating Characteristic Curves for Cure Survival Data and Mismeasured Biomarkers

Li-Pang Chen 

Department of Statistics, National Chengchi University, Taipei City 116, Taiwan; lchen723@nccu.edu.tw

**Abstract:** Cure models and receiver operating characteristic (ROC) curve estimation are two important issues in survival analysis and have received attention for many years. In the development of biostatistics, these two topics have been well discussed separately. However, a rare development in the estimation of the ROC curve has been made available based on survival data with the cure fraction. On the other hand, while a large body of estimation methods have been proposed, they rely on an implicit assumption that the variables are precisely measured. In applications, measurement errors are generally ubiquitous and ignoring measurement errors can cause unexpected bias for the estimator and lead to the wrong conclusion. In this paper, we study the estimation of the ROC curve and the area under curve (AUC) when variables or biomarkers are subject to measurement error. We propose a valid procedure to handle measurement error effects and estimate the parameters in the cure model, as well as the AUC. We also make an effort to establish the theoretical properties with rigorous justification.

**Keywords:** area under curve; bias correction; cure; censoring; EM algorithm; incomplete response; insertion method; mismeasurement

**MSC:** 62Nxx; 62N02



Academic Editor: Jinfeng Xu

Received: 13 December 2024

Revised: 20 January 2025

Accepted: 25 January 2025

Published: 27 January 2025

**Citation:** Chen, L.-P. Analysis of Receiver Operating Characteristic Curves for Cure Survival Data and Mismeasured Biomarkers.

*Mathematics* **2025**, *13*, 424. <https://doi.org/10.3390/math13030424>

**Copyright:** © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Survival analysis has been an important research topic in biostatistics. The main purpose of survival analysis is to understand the time at which a specific event or disease happens among patients and make a prediction. The key feature of survival data is that the time-to-event variable either is observed at a finite time (i.e., the event, such as death, happens during the observation period) or cannot be observed due to censoring. A more detailed introduction can be found in some monographs, such as [1]. However, in the context of survival analysis, there is a situation in which patients never experience the failure event (e.g., death) in the study period; this phenomenon is called cure, and the corresponding model and data structure is the well-known cure model. There are some estimation methods to analyze the cure model in the literature. For example, [2] considered the transformation model for survival data and implemented the logistic regression model to characterize the probability of cure. Ref. [3] proposed the conditional likelihood function given truncation variables. Ref. [4] explored the cure model with a mixture of single-index and Cox models. More comprehensive discussions are summarized in [5].

In applications, noisy data are usually inevitable due to the data collection or sampling mechanism. One characteristic of typical noisy data is the measurement error, which indicates that the observed variables cannot fully reflect what they should be. This phenomenon is usually caused by machines that are imprecise or incorrect records by researchers. In the

framework of survival data with cure models, some methods have also been developed to deal with measurement error problems. To name a few, ref. [6] proposed the simulation and extrapolation (SIMEX) method to correct the measurement errors. Ref. [7] developed the conditional expectation approach to correct the error effects. Unlike [6,7], who focused on the Cox proportional hazards model, ref. [8] studied the transformation model and adopted the SIMEX method to handle measurement error.

The other important issue in survival analysis is the estimation of the receiver operating characteristic (ROC) curve, whose primary goal is to display the sensitivity and the specificity of a continuous marker for a given disease. From the existing literature, the estimation procedures lie in the censored data, including [9–13]. However, to our knowledge, rare methods have been available to estimate the ROC curve when survival data contain the cured group and variables or biomarkers suffer from measurement error, except for [14,15], who considered the Cox model and the mixture cure model with the decomposition of sensitivity and specificity, respectively. The other challenging feature in analysis is the measurement error in biomarkers, which frequently appears in real-world applications. For example, [16] studied the Mayo Clinic primary biliary cirrhosis dataset with measurement errors in serum bilirubin, serum albumin, and prothrombin time; ref. [17] pointed out the biomarker systolic blood pressure (SBP), which is one of the key prognostic factors for cardiovascular risk scores, is possibly subject to measurement error; ref. [18] implemented three error-prone biomarkers, ER, PR, and Ki67 proteins, to analyze health study data on breast cancer in nurses. All the literature reveals that measurement error in biomarkers possibly affects the performance of standard diagnostic measures, such as ROC curves and area under the ROC curve (AUC). While error-prone biomarkers were discussed in survival analysis or the estimation of ROC curves (e.g., [16,17]), it seems that estimation methods for the cure fraction and error-prone biomarkers are unavailable.

To fill this research gap, we aim to develop a new estimation procedure in this paper. Specifically, we first consider the transformation model with the cure fraction in the survival data and the measurement error in covariates or biomarkers. Our setting generalizes the conventional implementation of the Cox model. Unlike existing methods that adopted the SIMEX method to correct for measurement error (e.g., [6,8]), we propose the insertion method, whose key spirit is to construct the new function so that its conditional expectation can be recovered to the true function based on the unobserved covariates or biomarkers. The corrected estimation function ensures the consistency of the estimator. After that, we further explore the estimation of the ROC curve when the survival data contain cured samples and mismeasured variables. We develop the time-independent and time-dependent estimation for the ROC curves and the corresponding estimation of the area under curve (AUC). Theoretically, we also establish the consistency and the asymptotic normality of the AUC estimator. The contribution of this paper is a valid estimation strategy, and the main focus is its theoretical establishment with rigorous justification.

The remainder is organized as follows. In Section 2, we introduce notation and models. In Section 3, we first present the proposed method to correct the measurement error effect and derive the estimator. Moreover, the corresponding theoretical properties are presented. In Section 4, we present numerical experiments and their results. We conclude the paper with discussion in Section 5. Finally, the proofs of the theoretical properties are placed in Appendix A.

## 2. Notation and Models

### 2.1. Cure Model

Let  $T$  and  $C$  denote the (uncured) failure time and the censoring time, respectively. With subjects cured and uncured, the failure time is determined by  $\tilde{T} \triangleq AT + (1 - A)\infty$ ,

where  $A \in \{0, 1\}$  indicates whether a subject is cured ( $A = 0$ ) or uncured ( $A = 1$ ). To characterize  $A$ , we consider the logistic regression

$$\pi(X) \triangleq P(A = 1|X) = \frac{\exp(\alpha^\top X)}{1 + \exp(\alpha^\top X)}, \quad (1)$$

where  $X$  is the  $p$ -dimensional vector of covariates or biomarkers and  $\alpha$  is the  $p$ -dimensional vector of the parameters associated with  $X$ .

On the other hand, regarding the uncured failure time  $T$ , we consider the survivor function that follows the transformation model (e.g., [19,20])

$$S_u(t|X_i) = \frac{\exp(-\beta^\top X_i/\rho)}{\{\exp(-\beta^\top X_i) + \rho H(t)\}^{1/\rho}}, \quad (2)$$

where  $\beta$  is the  $p$ -dimensional vector of the parameters associated with  $X$ ,  $H(t)$  is the unknown strictly increasing function, and  $\rho$  is the known parameter. In particular, when  $\rho = 1$ , (2) follows the proportional odds (PO) model; when  $\rho = 0$ , then (2) becomes the proportional hazards (PH) model. Given (2), the survivor function of  $\tilde{T}$  is given by

$$\begin{aligned} S(t|X) &\triangleq P(\tilde{T} > t|X) \\ &= P(\tilde{T} > t|X, A = 1)P(A = 1|X) + P(\tilde{T} > t|X, A = 0)P(A = 0|X) \\ &= P(T > t|X)\pi(X) + P(\infty > t|X)\{1 - \pi(X)\} \\ &= S_u(t|X)\pi(X) + 1 - \pi(X). \end{aligned}$$

Finally, let  $C$  denote the censoring time, and we further define  $Y \triangleq \min\{T, C\}$  as the observed survival time and  $\delta = \mathbb{I}(T \leq C)$ , with  $\mathbb{I}(\cdot)$  being the indicator function. Let  $\{Y_i, \delta_i, X_i\} : i = 1, \dots, n$  denote the independent and identically distributed sample as  $\{Y, \delta, X\}$ .

To develop the method in the following sections, some typical assumptions are imposed in the context of survival analysis, including noninformative censoring times, and the independence of failure time is independent of the censoring time, given the covariates.

## 2.2. Measurement Error Model

In applications, the covariates or biomarkers are possibly subject to measurement error. Here, we consider the case where  $X_i$  is error-contaminated, and let  $X_i^*$  denote the observed or surrogate measurement of  $X_i$ . Consistent with most work in the literature (e.g., [21]), we consider that the  $X_i$  are continuous and linked with  $X_i^*$  by the additive measurement error model:

$$X_i^* = X_i + \epsilon_i, \quad (3)$$

where  $\epsilon_i$  is independent of  $\{Y_i, \delta_i, X_i\}$ ,  $\epsilon_i \sim N(0_p, \Sigma_\epsilon)$  with the covariance matrix  $\Sigma_\epsilon$ , and  $0_p$  represents the  $p \times 1$  zero vector. Let  $\omega_{jk} = \frac{\text{cov}(X_{ij}, X_{ik})}{\text{cov}(X_{ij}^*, X_{ik}^*)}$  denote the *reliability ratio* of the covariances of  $j$ th and  $k$ th covariates and their surrogates for  $j, k = 1, \dots, p$ , which reflects the magnitude of variation between the unobserved and observed covariates, and higher values indicate a minor measurement error. Note that  $\text{cov}(X_{ij}^*, X_{ik}^*) = \text{cov}(X_{ij}, X_{ik}) + \text{cov}(\epsilon_{ij}, \epsilon_{ik})$ , which implies that  $\omega_{jk} = \frac{\text{cov}(X_{ij}, X_{ik})}{\text{cov}(X_{ij}, X_{ik}) + \text{cov}(\epsilon_{ij}, \epsilon_{ik})}$ . In applications, we may assume that the reliability ratio  $\omega_{jk}$  is not a small value for all  $j, k = 1, \dots, p$  (e.g., [22]), which implies that the values in  $\Sigma_\epsilon$  are not too large.

In situations where the parameters for the measurement error model (3) are unknown, we may utilize the information carried by additional data sources, such as repeated mea-

surements or validation subsamples. As a result, to highlight key ideas, we focus our attention on the estimation of parameters associated with the survival model and assume that the parameter  $\Sigma_\epsilon$  for measurement error model (3) is known. This assumption is reasonable, typically arising in two circumstances: (i) prior studies provide the information on the covariate mismeasurement and offer an estimate of  $\Sigma_\epsilon$ , and (ii) in conducting sensitivity analyses, different values of  $\Sigma_\epsilon$  are specified to understand how mismeasurement effects may affect inference results about the parameters associated with the survival model.

### 3. Methodology

Suppose that we have a sample of  $n$  subjects and that for  $i = 1, \dots, n$ ,  $\{Y_i, \delta_i, X_i^*\}$  has the same distribution as  $\{Y, \delta, X^*\}$ . We first estimate the unknown parameters  $(\alpha, \beta, H(\cdot))$  and the survivor function of  $\tilde{T}$  based on formulation (2) with suitable error correction. After that, we develop the error-corrected method to estimate the time-independent/time-dependent AUC. Moreover, theoretical results are also established after presenting the proposed estimation procedures.

#### 3.1. Construction of the Error-Corrected Likelihood Function

To develop our approach, we start the discussion by pretending that  $X_i$  is available. After that, we replace  $X_i$  with  $X_i^*$  and propose the error-corrected method to adjust error effects.

For cured patients with  $\delta_i = 0$  and  $A_i = 0$ , we have  $P(\tilde{T}_i = \infty)^{1-A_i} = \{1 - \pi(X_i)\}^{1-A_i}$ . In addition, for those patients who encounter the failure event, i.e.,  $\delta_i = 1$  and  $A_i = 1$ , we have  $\{f_T(y|X_i)\pi(X_i)\}^{\delta_i=1}$ , where  $f_T(\cdot)$  represents the density function of uncured failure time. Moreover, for observations who are censored but not cured ( $\delta_i = 0$  and  $A_i = 1$ ), we can obtain  $\{S_u(y|X_i)\pi(X_i)\}^{\mathbb{I}(\delta_i=0, A_i=1)}$ . Therefore, for subject  $i = 1, \dots, n$ , the complete likelihood function is given by

$$L \triangleq \prod_{i=1}^n \{1 - \pi(X_i)\}^{1-A_i} \{\pi(X_i)\}^{A_i} \times \left\{ \frac{dH(Y_i)}{\exp(-\beta^\top X_i) + \rho H(Y_i)} \right\}^{\delta_i} \left\{ \frac{\exp(-\beta^\top X_i)}{\exp(-\beta^\top X_i) + \rho H(Y_i)} \right\}^{A_i/\rho}, \quad (4)$$

where  $dH(\cdot)$  is the derivative of  $H(\cdot)$ , and the corresponding log-likelihood function of (4) is

$$\begin{aligned} \ell = & \sum_{i=1}^n \left( A_i X_i^\top \alpha - \log \{1 + \exp(X_i^\top \alpha)\} \right. \\ & + \delta_i \left[ \log \{dH(Y_i)\} - \log \{ \exp(-\beta^\top X_i) + \rho H(Y_i) \} \right] \\ & \left. - \frac{A_i}{\rho} \left[ \beta^\top X_i + \log \{ \exp(-\beta^\top X_i) + \rho H(Y_i) \} \right] \right). \end{aligned} \quad (5)$$

However, note that we usually have  $X_i^*$  in (5) instead of  $X_i$ . It is crucial to make a suitable correction. Our strategy is to find the new likelihood function based on  $X_i^*$ , denoted by  $\ell^*$ , such that the conditional expectation  $E(\ell^*|X_i) = \ell$  is satisfied. In this way, we have  $E(\ell^*) = E(\ell)$ , showing that  $\ell^*$  is an error-corrected likelihood function and provides the same optimizer as those obtained from  $\ell$ . By measurement error model (3) and the moment-generating function under the normal distribution, it is direct to have  $E(X_i^*|X_i) = X_i$  and  $E\{\exp(\alpha^\top X_i^*)|X_i\} = \exp(\alpha^\top X_i + \frac{1}{2}\alpha^\top \Sigma_\epsilon \alpha)$ . According to (5), however, the challenge is that the error-prone covariate appears in  $\log\{1 + \exp(X_i^{*\top} \alpha)\}$  and  $\log\{\exp(-\beta^\top X_i^*) + \rho H(Y_i)\}$ , which are nonlinear terms.

To address this issue, we first consider the function  $\frac{X_i^* \exp(\alpha^\top X_i^*)}{1 + \exp(\alpha^\top X_i^*)}$ . Since the expectation of a ratio random variable is sometimes difficult to compute, we may adopt the approximation. Specifically, let  $\Xi_{1i} \triangleq X_i^* \exp(\alpha^\top X_i^*)$  and  $\Xi_{2i} \triangleq 1 + \exp(\alpha^\top X_i^*)$ . By the second-order Taylor series expansion (e.g., [23], pp. 69–72), we have the following approximation:

$$\begin{aligned} \frac{\Xi_{1i}}{\Xi_{2i}} &\approx \frac{E(\Xi_{1i}|X_i)}{E(\Xi_{2i}|X_i)} + \frac{1}{E(\Xi_{2i}|X_i)} \{\Xi_{1i} - E(\Xi_{1i}|X_i)\} - \frac{E(\Xi_{1i}|X_i)}{\{E(\Xi_{2i}|X_i)\}^2} \{\Xi_{2i} - E(\Xi_{2i}|X_i)\} \\ &\quad - \frac{1}{2} \frac{1}{\{E(\Xi_{2i}|X_i)\}^2} \{\Xi_{1i} - E(\Xi_{1i}|X_i)\} \{\Xi_{2i} - E(\Xi_{2i}|X_i)\} \\ &\quad + \frac{1}{2} \frac{E(\Xi_{1i}|X_i)}{\{E(\Xi_{2i}|X_i)\}^3} \{\Xi_{2i} - E(\Xi_{2i}|X_i)\}^2. \end{aligned}$$

Taking the conditional expectation gives that

$$\begin{aligned} E \left\{ \frac{X_i^* \exp(\alpha^\top X_i^*)}{1 + \exp(\alpha^\top X_i^*)} \middle| X_i \right\} &\approx \frac{(X + \Sigma_\epsilon \alpha) \exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha)}{1 + \exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha)} \\ &\quad - \frac{1}{2} \frac{\text{cov}(\Xi_{1i}, \Xi_{2i}|X_i)}{\left\{1 + \exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha)\right\}^2} \\ &\quad + \frac{1}{2} \frac{(X + \Sigma_\epsilon \alpha) \exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha)}{\left\{1 + \exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha)\right\}^3} \text{var}(\Xi_{2i}|X_i). \end{aligned} \quad (6)$$

On the other hand, the conditional variance of  $\Xi_{2i}$  and conditional covariance of  $\Xi_{1i}$  and  $\Xi_{2i}$ , given  $X_i$ , are respectively derived as

$$\begin{aligned} \text{var}(\Xi_{2i}|X_i) &= \text{var}\{1 + \exp(\alpha^\top X_i^*)|X_i\} \\ &= \text{var}\{\exp(\alpha^\top X_i^*)|X_i\} \\ &= E\{\exp(2\alpha^\top X_i^*)|X_i\} - [E\{\exp(\alpha^\top X_i^*)|X_i\}]^2 \\ &= \exp(2\alpha^\top X_i + 2\alpha^\top \Sigma_\epsilon \alpha) - \exp(2\alpha^\top X_i + \alpha^\top \Sigma_\epsilon \alpha) \end{aligned}$$

and

$$\begin{aligned} \text{cov}(\Xi_{1i}, \Xi_{2i}|X_i) &= E\{X_i^* \exp(\alpha^\top X_i^*) + X_i^* \exp(2\alpha^\top X_i^*)|X_i\} \\ &\quad - E\{X_i^* \exp(\alpha^\top X_i^*)|X_i\} E\{1 + X_i^* \exp(\alpha^\top X_i^*)|X_i\} \\ &= (X_i + 2\Sigma_\epsilon \alpha) \exp(2\alpha^\top X_i + 2\alpha^\top \Sigma_\epsilon \alpha) - (X_i + \Sigma_\epsilon \alpha) \exp(2\alpha^\top X_i + \alpha^\top \Sigma_\epsilon \alpha). \end{aligned}$$

Provided that the reliability ratio is not too small or, equivalently, that the variance of the noise term  $\Sigma_\epsilon$  is not too large, the two values  $\text{var}(\Xi_{2i}|X_i)$  and  $\text{cov}(\Xi_{1i}, \Xi_{2i}|X_i)$  are close to zero, yielding that the second and last terms in (6) are close to zero; thus, we have that

$$E \left\{ \frac{X_i^* \exp(\alpha^\top X_i^*)}{1 + \exp(\alpha^\top X_i^*)} \middle| X_i \right\} \approx \frac{(X + \Sigma_\epsilon \alpha) \exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha)}{1 + \exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha)}, \quad (7)$$

and performing integration on (7) with respect to  $\alpha$  yields

$$\begin{aligned} E \left[ \log \left\{ 1 + \exp(\alpha^\top X_i^*) \right\} \middle| X_i \right] &\approx \log \left\{ 1 + \exp \left( \alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha \right) \right\} \\ &= \log \left\{ 1 + \exp(\alpha^\top X_i) \right\} + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha, \end{aligned} \quad (8)$$

where the last approximation is due to that “plus one” being almost ignorable. Thus, (8) suggests that the suitable error correction of  $\log\{1 + \exp(\alpha^\top X_i)\}$  is

$$\log\{1 + \exp(\alpha^\top X_i^*)\} - \frac{1}{2}\alpha^\top \Sigma_\epsilon \alpha. \quad (9)$$

Further, similar derivations of (7) and (8) enable us to obtain the following equality:

$$\begin{aligned} & E\left[\log\{\rho H(Y_i) + \exp(-\beta^\top X_i^*)\} \middle| X_i\right] \\ &= \log\left\{\rho H(Y_i) + \exp\left(-\beta^\top X_i + \frac{1}{2}\beta^\top \Sigma_\epsilon \beta\right)\right\}, \end{aligned}$$

yielding that the suitable error correction of  $\log\{\rho H(Y_i) + \exp(-\beta^\top X_i)\}$  is

$$\log\{\rho H(Y_i) + \exp(-\beta^\top X_i^*)\} - \frac{1}{2}\beta^\top \Sigma_\epsilon \beta. \quad (10)$$

Consequently, combining (5) with (9) and (10) yields the corrected log-likelihood function

$$\begin{aligned} \ell^* &= \sum_{i=1}^n \left( A_i \alpha^\top X_i^* - \log\{1 + \exp(\alpha^\top X_i^*)\} + \frac{1}{2}\alpha^\top \Sigma_\epsilon \alpha \right. \\ &\quad \left. + \delta_i \left[ \log\{dH(Y_i)\} - \log\{\rho H(Y_i) + \exp(-\beta^\top X_i^*)\} + \frac{1}{2}\beta^\top \Sigma_\epsilon \beta \right] \right. \\ &\quad \left. - \frac{A_i}{\rho} \left[ \beta^\top X_i^* + \log\{\rho H(Y_i) + \exp(-\beta^\top X_i^*)\} - \frac{1}{2}\beta^\top \Sigma_\epsilon \beta \right] \right). \end{aligned} \quad (11)$$

### 3.2. Estimation of Parameters and Functions

After establishing the corrected log-likelihood function (11), we aim to estimate unknown parameters  $\alpha$  and  $\beta$ , as well as unknown function  $H(\cdot)$ .

However, the other concern is that for those censored observations ( $\delta_i = 0$ ), whether they are cured or not is unknown; thus,  $A_i$  can be regarded as a missing value. To address this concern, we apply the expectation–maximization (EM) algorithm with error correction. Specifically, in the E-step, we start by considering  $E(A_i|\delta_i, \tilde{T}_i, X_i)$  to replace  $A_i$ . Recall that we have  $X_i^*$  only instead of  $X_i$ ; then, it is crucial to make a suitable adjustment. Let  $\xi_i \triangleq E(A_i|\delta_i, \tilde{T}_i, X_i)$ . We aim to find  $\xi_i^* \triangleq E(\xi_i|X_i^*)$ , which is in terms of the observed covariate  $X_i$ , such that the measurement error would be adjusted, i.e.,  $E(\xi_i^*) = E(\xi_i)$ .

When  $\delta_i = 1$ , we have  $E(A_i|\delta_i = 1, \tilde{T}_i, X_i) = 1$ , and it is easy to check that  $\xi_i^* = 1$ . When  $\delta_i = 0$ , then  $E(A_i|\delta_i = 0, \tilde{T}_i, X_i) = \frac{\pi(X_i)S_u(Y_i|X_i)}{1-\pi(X_i)+\pi(\tilde{X}_i)S_u(Y_i|\tilde{X}_i)}$ . Let  $\tilde{X}_i \triangleq E(X_i|X_i^*)$ , which satisfies  $E(\tilde{X}_i) = E(X_i)$ . In particular, we apply the idea in [22] to express  $\tilde{X}_i$  by the linear approximation

$$\tilde{X}_i \triangleq \mu_X + \Sigma_X^{1/2} \Sigma_{X^*}^{-1/2} (X_i^* - \mu_{X^*}) \quad (12)$$

with  $\mu_X \triangleq E(X)$ ,  $\mu_{X^*} \triangleq E(X^*)$ ,  $\Sigma_X \triangleq \text{var}(X)$ , and  $\Sigma_{X^*} \triangleq \text{var}(X^*)$ .

Following the derivation of [22], we apply the second-order Taylor series expansion for  $\xi_i$  and  $\xi_i^* \triangleq \frac{\pi(\tilde{X}_i)S_u(Y_i|\tilde{X}_i)}{1-\pi(\tilde{X}_i)+\pi(\tilde{X}_i)S_u(Y_i|\tilde{X}_i)}$  around  $\mu_X$  with respect to  $X_i$  and  $\tilde{X}_i$ , respectively; then, we can obtain the following two approximations:

$$\begin{aligned}\xi_i &\approx \frac{\pi(\mu_X)S_u(Y_i|\mu_X)}{1 - \pi(\mu_X) + \pi(\mu_X)S_u(Y_i|\mu_X)} + \left( \frac{\partial \xi_i}{\partial X_i} \bigg|_{X_i=\mu_X} \right)^\top (X_i - \mu_X) \\ &\quad + \frac{1}{2}(X_i - \mu_X)^\top \frac{\partial^2 \xi_i}{\partial X_i \partial X_i^\top} \bigg|_{X_i=\mu_X} (X_i - \mu_X)\end{aligned}\quad (13)$$

and

$$\begin{aligned}\tilde{\xi}_i &\approx \frac{\pi(\mu_X)S_u(Y_i|\mu_X)}{1 - \pi(\mu_X) + \pi(\mu_X)S_u(Y_i|\mu_X)} + \left( \frac{\partial \tilde{\xi}_i}{\partial \tilde{X}_i} \bigg|_{\tilde{X}_i=\mu_X} \right)^\top (\tilde{X}_i - \mu_X) \\ &\quad + \frac{1}{2}(\tilde{X}_i - \mu_X)^\top \frac{\partial^2 \tilde{\xi}_i}{\partial \tilde{X}_i \partial \tilde{X}_i^\top} \bigg|_{\tilde{X}_i=\mu_X} (\tilde{X}_i - \mu_X).\end{aligned}\quad (14)$$

Combining (13) and (14) gives that

$$\begin{aligned}\xi_i &\approx \tilde{\xi}_i + \left( \frac{\partial \xi_i}{\partial X_i} \bigg|_{X_i=\mu_X} \right)^\top (X_i - \tilde{X}_i) + \frac{1}{2}(X_i - \mu_X)^\top \frac{\partial^2 \xi_i}{\partial X_i \partial X_i^\top} \bigg|_{X_i=\mu_X} (X_i - \mu_X) \\ &\quad - \frac{1}{2}(\tilde{X}_i - \mu_X)^\top \frac{\partial^2 \tilde{\xi}_i}{\partial \tilde{X}_i \partial \tilde{X}_i^\top} \bigg|_{\tilde{X}_i=\mu_X} (\tilde{X}_i - \mu_X).\end{aligned}\quad (15)$$

Then, taking the expectation on (15) yields that

$$E(\xi_i) \approx E(\tilde{\xi}_i), \quad (16)$$

where the expectation of the subtraction of two quadratic forms in (16) is given by

$$\begin{aligned}&E \left\{ (X_i - \mu_X)^\top \frac{\partial^2 \xi_i}{\partial X_i \partial X_i^\top} \bigg|_{X_i=\mu_X} (X_i - \mu_X) - (\tilde{X}_i - \mu_X)^\top \frac{\partial^2 \tilde{\xi}_i}{\partial \tilde{X}_i \partial \tilde{X}_i^\top} \bigg|_{\tilde{X}_i=\mu_X} (\tilde{X}_i - \mu_X) \right\} \\ &= \text{trace} \left( \frac{\partial^2 \xi_i}{\partial X_i \partial X_i^\top} \bigg|_{X_i=\mu_X} \text{var}(X_i) \right) - \text{trace} \left( \frac{\partial^2 \tilde{\xi}_i}{\partial \tilde{X}_i \partial \tilde{X}_i^\top} \bigg|_{\tilde{X}_i=\mu_X} \text{var}(\tilde{X}_i) \right) \\ &= 0\end{aligned}$$

since  $\text{var}(\tilde{X}_i) = \Sigma_X^{1/2} \Sigma_{X^*}^{-1/2} \text{var}(X_i^*) \Sigma_{X^*}^{-1/2} \Sigma_X^{1/2} = \Sigma_X = \text{var}(X_i)$ . Consequently, (16) implies that

$$\xi_i^* = E(\xi_i|X_i^*) \approx \frac{\pi(\tilde{X}_i)S_u(Y_i|\tilde{X}_i)}{1 - \pi(\tilde{X}_i) + \pi(\tilde{X}_i)S_u(Y_i|\tilde{X}_i)}. \quad (17)$$

Therefore, by the given data and the fact  $\Sigma_{X^*} = \Sigma_X + \Sigma_\epsilon$  from measurement error model (3), (12) can be estimated by

$$\hat{X}_i \triangleq \hat{\mu}_{X^*} + \left( \hat{\Sigma}_{X^*} - \Sigma_\epsilon \right)^{1/2} \hat{\Sigma}_{X^*}^{-1/2} (X_i^* - \hat{\mu}_{X^*}),$$

where  $\hat{\mu}_{X^*}$  and  $\hat{\Sigma}_{X^*}$  are the empirical estimates of  $\mu_{X^*}$  and  $\Sigma_{X^*}$ , respectively. Thus, replacing  $\tilde{X}_i$  in (17) with  $\hat{X}_i$  yields

$$\hat{\xi}_i^* = \frac{\pi(\hat{X}_i)S_u(Y_i|\hat{X}_i)}{1 - \pi(\hat{X}_i) + \pi(\hat{X}_i)S_u(Y_i|\hat{X}_i)}. \quad (18)$$

When the E-step is established, we examine the M-step. Specifically, replacing  $A_i$  in (11) with  $\zeta_i^*$  gives

$$\begin{aligned}\widehat{\ell}^* &= \sum_{i=1}^n \left( \zeta_i^* \alpha^\top X_i^* - \log \{ 1 + \exp(\alpha^\top X_i^*) \} + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha \right) \\ &\quad + \sum_{i=1}^n \left( \delta_i \left[ \log \{ dH(Y_i) \} - \log \{ \rho H(Y_i) + \exp(-\beta^\top X_i^*) \} + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta \right] \right. \\ &\quad \left. - \frac{\zeta_i^*}{\rho} \left[ \beta^\top X_i^* + \log \{ \rho H(Y_i) + \exp(-\beta^\top X_i^*) \} - \frac{1}{2} \beta^\top \Sigma_\epsilon \beta \right] \right) \\ &\triangleq \widehat{\ell}_1^* + \widehat{\ell}_2^*,\end{aligned}\quad (19)$$

where  $\widehat{\ell}_1^*$  contains information on  $\alpha$  only and  $\widehat{\ell}_2^*$  includes  $\beta$  and  $H(\cdot)$ , which reflect the survival process. Thus, the maximization of  $\widehat{\ell}^*$  for  $(\alpha, \beta, H(\cdot))$  can be performed separately. More specifically, the estimator of  $\alpha$ , denoted by  $\widehat{\alpha}$ , is determined by  $\widehat{\alpha} = \underset{\alpha}{\operatorname{argmax}} \widehat{\ell}_1^*$ . Due to the unavailability of the closed form, one can adopt the Newton–Raphson approach to derive  $\widehat{\alpha}$ .

Naturally, the estimator of  $(\beta, H(\cdot))$  can also be determined by

$$(\widehat{\beta}, \widehat{H}(\cdot)) = \underset{(\beta, H(\cdot))}{\operatorname{argmax}} \widehat{\ell}_2^*. \quad (20)$$

However, the challenge comes from nonparametric function  $H(\cdot)$ . To address it, we follow the same line as [19] and “parametrize” function  $H(\cdot)$  by  $\sum_{j=1}^{d_i} \exp(\gamma_j)$ , where  $d_i \triangleq \max\{j : t_{(j)} \leq t_i\}$  represents the index of the latest jump point before  $t_i$  and  $\gamma_1, \dots, \gamma_m$  are the corresponding logs of jump sizes of  $H(\cdot)$  at  $t_{(1)}, \dots, t_{(m)}$ , which are the sorted times. Based on such representation,  $\widehat{\ell}_2^*$  can be expressed as

$$\begin{aligned}\widehat{\ell}_2^* &= \sum_{i=1}^n \left( \delta_i \left[ \gamma_i - \log \left\{ \rho \sum_{j=1}^{d_i} \exp(\gamma_j) + \exp(-\beta^\top X_i^*) \right\} + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta \right] \right. \\ &\quad \left. - \frac{\zeta_i^*}{\rho} \left[ \beta^\top X_i^* + \log \left\{ \rho \sum_{j=1}^{d_i} \exp(\gamma_j) + \exp(-\beta^\top X_i^*) \right\} - \frac{1}{2} \beta^\top \Sigma_\epsilon \beta \right] \right). \quad (21)\end{aligned}$$

Note that (21) is a concave function; then, taking the negative sign on (21) makes it become convex. To derive the estimators of  $\beta$  and  $\gamma \triangleq (\gamma_1, \dots, \gamma_m)$  from (21), we employ the coordinate-descent algorithm, which is frequently implemented in the optimization problem (e.g., [24]). Detailed descriptions of the steps are given below:

Step 1: Choose an initial value for  $\beta$  and  $\gamma$ , and denote them by  $\beta^{(0)}$  and  $\gamma^{(0)}$ .

Step 2: For  $r = 1, \dots, p$ , given  $\gamma^{(k-1)} \triangleq (\gamma_1^{(k-1)}, \dots, \gamma_p^{(k-1)})$  and  $\beta_1^{(k-1)}, \dots, \beta_{r-1}^{(k-1)}, \beta_{r+1}^{(k-1)}, \dots, \beta_p^{(k-1)}$ , update  $\beta_r$  by finding

$$\beta_r^{(k)} = \underset{\beta_r}{\operatorname{argmin}} -\widehat{\ell}_2^*(\beta_r | \gamma^{(k-1)}, \beta_1^{(k-1)}, \dots, \beta_{r-1}^{(k-1)}, \beta_{r+1}^{(k-1)}, \dots, \beta_p^{(k-1)}).$$

Step 3: For  $s = 1, \dots, m$ , given  $\beta^{(k)} \triangleq (\beta_1^{(k)}, \dots, \beta_p^{(k)})$  and  $\gamma_1^{(k-1)}, \dots, \gamma_{r-1}^{(k-1)}, \gamma_{r+1}^{(k-1)}, \dots, \gamma_p^{(k-1)}$ , update  $\gamma_s$  by finding

$$\gamma_s^{(k)} = \underset{\gamma_r}{\operatorname{argmin}} -\widehat{\ell}_2^*(\gamma_r | \beta^{(k)}, \gamma_1^{(k-1)}, \dots, \gamma_{r-1}^{(k-1)}, \gamma_{r+1}^{(k-1)}, \dots, \gamma_m^{(k-1)}).$$

Step 4: Repeat Steps 2 and 3 until convergence, and let  $\widehat{\beta}$  and  $\widehat{\gamma}$  denote the limit of  $\beta^{(k)}$  and  $\gamma^{(k)}$  as  $k \rightarrow \infty$ .

We comment that the coordinate-descent algorithm is a common strategy to obtain the minimizer (e.g., [24]). Its key idea is to iteratively update the scale parameter with other values given under the previous step. This approach enables us to obtain a stable numeric scheme and avoid estimating high-dimensional vectors of parameters. In the literature, the alternative strategy to address the optimization in (20) is called the minimization–maximization (MM) algorithm (e.g., [19,25]). While the MM algorithm is also an iterative approach based on the historical values that is similar to the coordinate-descent algorithm, the key step of this approach is to find and optimize the surrogate function that improves the value of the objective function or leave it unchanged, but in general, such suitable surrogate functions are not easy to find. Instead, the coordinate-descent algorithm enables us to directly examine (21).

Finally, to assess the validity of the proposed method, we present the following theoretical results, and the proof is postponed to Appendix A.2.

**Theorem 1.** Let  $\alpha_0$ ,  $\beta_0$ , and  $H_0(\cdot)$  denote the true values of parameters. Under regularity conditions, the proposed estimator is consistent. That is, as  $n \rightarrow \infty$ ,

$$\|\widehat{\alpha} - \alpha_0\|_2 \xrightarrow{p} 0, \quad \|\widehat{\beta} - \beta_0\|_2 \xrightarrow{p} 0, \quad \text{and} \quad \|\widehat{H}(t) - H_0(t)\|_\infty \xrightarrow{p} 0,$$

where  $\|\cdot\|_2$  and  $\|\cdot\|_\infty$  are the  $L_2$ - and supremum norms, respectively.

Let  $\mathcal{V} = \{v = (v_1, v_2, v_3)\}$ , and let us define  $\mathcal{V}_p = \{v \in \mathcal{V} : \|v_1\|, \|v_2\|, \|v_3\|_{\text{BV}} < p\}$  as the collection of directions, where  $v_1$  and  $v_2$  are  $p$ -dimensional vectors,  $v_3$  is a function defined in  $[0, \tau]$  with  $\tau$  being the end time of the study, and  $\|\cdot\|_{\text{BV}}$  is the total variation, defined as

$$\|v_3\|_{\text{BV}} = \sup_{0=t_0 < t_1 < \dots < t_n=\tau} \sum_{j=1}^n |v_3(t_j) - v_3(t_{j-1})|.$$

The next theorem shows the convergence in distribution of the proposed method with the rate  $\sqrt{n}$ .

**Theorem 2.** Under regularity conditions, the proposed estimator has an asymptotic distribution. That is, as  $n \rightarrow \infty$ ,  $\sqrt{n}(\widehat{\alpha} - \alpha_0, \widehat{\beta} - \beta_0, \widehat{H}(t) - H_0(t))$  converges in distribution to a mean-zero Gaussian process in the functional space  $l_\infty(\mathcal{V}_p)$  on  $\mathcal{V}_p$ .

Let  $\phi \triangleq (\alpha^\top, \beta^\top)^\top$ . According to (19), the parameters  $\alpha$  and  $\beta$  appear in  $\widehat{\ell}_1^*$  and  $\widehat{\ell}_2^*$ . With  $H(\cdot)$  fixed, by the theory of  $M$ -estimation, the marginal asymptotic distribution of  $\widehat{\phi} \triangleq (\widehat{\alpha}^\top, \widehat{\beta}^\top)^\top$  is given by

$$\sqrt{n}(\widehat{\phi} - \phi_0) \xrightarrow{d} N(0, \Sigma_\phi), \quad (22)$$

where  $\Sigma_\phi = \mathcal{I}^{-1} \mathcal{J} \mathcal{I}^{-1}$ , with

$$\mathcal{I} = \text{diag} \left( \frac{\partial^2}{\partial \alpha \partial \alpha^\top} \hat{\ell}_1^*, \frac{\partial^2}{\partial \beta \partial \beta^\top} \hat{\ell}_2^* \right), \quad \mathcal{J} = E(a^{\otimes 2}),$$

and  $a^{\otimes 2} = aa^\top$  for a vector  $a \triangleq \left( \left( \frac{\partial \hat{\ell}_1^*}{\partial \alpha} \right)^\top, \left( \frac{\partial \hat{\ell}_2^*}{\partial \beta} \right)^\top \right)^\top$ .

### 3.3. Estimation of ROC and AUC

When the estimators of  $H(\cdot)$ ,  $\alpha$ , and  $\beta$  are obtained, we now propose the estimation procedure for the time-independent and time-dependent ROC curves, as well as the area under curve (AUC) with measurement error, where the former says that the event status and marker value for an individual are independent of the change in the time and the latter reflects that the curve varies at different time points and seems more flexible than the former. In real-world applications, it is common for the disease status to change over time, so the time-dependent ROC curve is more widely explored. However, in the presence of the cure fraction in survival time and measurement error in the covariates, both time-independent and time-dependent cases may not be fully discussed. To make the discussion on measurement error effects more comprehensive, we still study two different types of ROC curve estimation in the following two subsections.

#### 3.3.1. Time-Independent AUC

We first discuss the time-independent case. Let  $C_1(X)$  be the binary and time-independent classifier with biomarker  $X$ , which classifies patients into the uncured group with 1. On the other hand, as discussed by [26], the linear composite biomarker is an optimal classifier among all functions of biomarkers. As a result, the event  $C_1(X) = 1$  is equivalent to the inequality  $\beta^\top X > u$  for some constant  $u$ . Then, the time-independent true-positive rate (TPR) and false-positive rate (FPR), denoted by  $\mathcal{T}(\alpha, \beta, u)$  and  $\mathcal{F}(\alpha, \beta, u)$ , are defined as

$$\mathcal{T}(\alpha, \beta, u) = P(\beta^\top X > u | \tilde{T} < \infty) \quad \text{and} \quad \mathcal{F}(\alpha, \beta, u) = P(\beta^\top X > u | \tilde{T} = \infty). \quad (23)$$

Moreover, (23) can be further expressed as

$$\begin{aligned} \mathcal{T}(\alpha, \beta, u) &= \frac{P(\beta^\top X > u, \tilde{T} < \infty)}{P(\tilde{T} < \infty)} \\ &= \frac{\int P(\tilde{T} < \infty | X = x) \mathbb{I}(\beta^\top X > u) f_X(x) dx}{\int P(\tilde{T} < \infty | X = x) f_X(x) dx} \\ &= \frac{E\{\pi(X) \mathbb{I}(\beta^\top X > u)\}}{E\{\pi(X)\}} \end{aligned} \quad (24)$$

and

$$\begin{aligned} \mathcal{F}(\alpha, \beta, u) &= \frac{P(\beta^\top X > u, \tilde{T} = \infty)}{P(\tilde{T} = \infty)} \\ &= \frac{\int P(\tilde{T} = \infty | X = x) \mathbb{I}(\beta^\top X > u) f_X(x) dx}{\int P(\tilde{T} = \infty | X = x) f_X(x) dx} \\ &= \frac{E[\{1 - \pi(X)\} \mathbb{I}(\beta^\top X > u)]}{E\{1 - \pi(X)\}} \end{aligned} \quad (25)$$

respectively, with  $f_X(\cdot)$  and  $E(\cdot)$  being the density function of  $X$  and the expectation evaluated with respect to  $X$ , respectively, where the last equality in (24) comes from

$$\begin{aligned} P(\tilde{T} < \infty | X = x) &= P(\tilde{T} < \infty | A = 1, X = x)P(A = 1 | X = x) \\ &= P(T < \infty | X = x)\pi(X) \\ &= \pi(X) \end{aligned}$$

with  $P(T < \infty | X = x) = \int_0^\infty f(t|x)dt = 1$  and  $f(t|x)$  being the conditional density function of  $T$  given  $X$ ; similar derivations show that

$$\begin{aligned} P(\tilde{T} = \infty | X = x) &= P(\tilde{T} = \infty | A = 0, X = x)P(A = 0 | X = x) \\ &= 1 - \pi(X) \end{aligned}$$

with  $P(\tilde{T} = \infty | A = 0, X = x) = 1$ , yielding the last equality of (25).

Therefore, based on (24) and (25), the AUC, denoted by  $\mathcal{A}(\alpha, \beta)$ , is given by (e.g., [11])

$$\mathcal{A}(\alpha, \beta) = \frac{E[\pi(X_1)\{1 - \pi(X_2)\}\mathbb{I}(\beta^\top X_1 > \beta^\top X_2)]}{E\{\pi(X)\}E\{1 - \pi(X)\}}. \quad (26)$$

In practice, however, biomarker  $X$  may also be measured with error. Let  $X^*$  denote the surrogate version of  $X$ . We adopt measurement error model (3) to build the relationship between  $X$  and  $X^*$ .

In (24),  $X^*$  appears in  $\pi(X)$  and  $\pi(X)\mathbb{I}(\beta^\top X > u)$ . To correct for error effects, our strategy is to follow a similar approach to that in Section 3.2 and find the surrogate functions in terms of  $X^*$ , say  $\pi^*(X^*)$  and  $\varphi^*(X^*)$ , such that  $\pi^*(X^*) = E\{\pi(X)|X^*\}$  and  $\varphi^*(X^*) = E\{\pi(X)\mathbb{I}(\beta^\top X > u)|X^*\}$ . In this way, we have  $E\{\pi^*(X^*)\} = E\{\pi(X)\}$  and  $E\{\varphi^*(X^*)\} = E\{\pi(X)\mathbb{I}(\beta^\top X > u)\}$ ; thus,  $\pi(X)$  and  $\pi(X)\mathbb{I}(\beta^\top X > u)$  can be replaced by  $\pi^*(X^*)$  and  $\varphi^*(X^*)$ , respectively. A similar strategy is also adopted for the numerator and denominator terms in (25) and (26). To this end, we develop  $\pi^*(X^*)$  and  $\varphi^*(X^*)$ .

According to the Mean Value Theorem, there exists  $X'$  between  $X$  and  $\tilde{X} \triangleq E(X|X^*)$ , such that

$$\pi(X) = \pi(\tilde{X}) + \frac{\partial \pi(X)}{\partial X} \Big|_{X=X'} (X - \tilde{X}), \quad (27)$$

and taking the conditional expectation  $E(\cdot|X^*)$  gives

$$E\{\pi(X)|X^*\} = \pi(\tilde{X}). \quad (28)$$

This indicates that the corrected version  $\pi^*(X^*)$  is given by the original function  $\pi(\cdot)$  with  $X$  being replaced by  $\tilde{X}$ . In particular, this result is essentially similar to the result in [27], which pointed out the satisfactory approximation of implementing  $\tilde{X}$  in the logit function.

On the other hand, note that based on the measurement error model, if  $\beta^\top X > u$ , then we have  $\beta^\top X^* > \beta^\top X > u - \|\beta\|\|\epsilon\|$ . This indicates that  $\mathbb{I}(\beta^\top X > u)$  implies  $\mathbb{I}(\beta^\top X^* > u)$ . Therefore,  $\pi(X)\mathbb{I}(\beta^\top X > u)$  takes the value  $\pi(X)$  when  $\mathbb{I}(\beta^\top X > u) = 1$ , and by a similar strategy to (27), taking the conditional expectation yields

$$E\{\pi(X)\mathbb{I}(\beta^\top X > u)|X^*\} = \pi(\tilde{X})\mathbb{I}(\beta^\top \tilde{X} > u) \quad (29)$$

thus,  $\varphi^*(X^*)$  is taken as  $\pi(X)\mathbb{I}(\beta^\top X > u)$  with  $X$  being replaced by  $\tilde{X}$ .

Consequently, the corrected TPR, FPR, and AUC are given by (24), (25), and (26), respectively. Moreover, with the consistent estimator  $\hat{\alpha}$  in Section 3.2, the corresponding estimators of the corrected TPR, FPR, and AUC are

$$\hat{\mathcal{T}}^*(\hat{\alpha}, \hat{\beta}, u) = \frac{\frac{1}{n} \sum_{i=1}^n \hat{\pi}(\tilde{X}_i) \mathbb{I}(\hat{\beta}^\top \tilde{X}_i > u)}{\frac{1}{n} \sum_{i=1}^n \hat{\pi}(\tilde{X}_i)}, \quad \hat{\mathcal{F}}^*(\hat{\alpha}, \hat{\beta}, u) = \frac{\frac{1}{n} \sum_{i=1}^n \{1 - \hat{\pi}(\tilde{X}_i)\} \mathbb{I}(\hat{\beta}^\top \tilde{X}_i > u)}{\frac{1}{n} \sum_{i=1}^n \{1 - \hat{\pi}(\tilde{X}_i)\}},$$

and

$$\hat{\mathcal{A}}(\hat{\alpha}, \hat{\beta}) = \frac{\frac{1}{n(n-1)} \sum_{i \neq j} [\hat{\pi}(\tilde{X}_i) \{1 - \hat{\pi}(\tilde{X}_j)\} \mathbb{I}(\hat{\beta}^\top \tilde{X}_i > \hat{\beta}^\top \tilde{X}_j)]}{\left\{ \frac{1}{n} \sum_{i=1}^n \hat{\pi}(\tilde{X}_i) \right\} \left[ \frac{1}{n} \sum_{i=1}^n \{1 - \hat{\pi}(\tilde{X}_i)\} \right]}, \quad (30)$$

where  $\hat{\pi}(\cdot)$  represents  $\pi(\cdot)$  with  $\alpha$  being replaced by  $\hat{\alpha}$ .

### 3.3.2. Time-Dependent AUC

We now explore the time-dependent case. Denote the time-dependent TPR and FPR, respectively, by

$$\mathcal{T}_t(\alpha, \beta, u) = P(\beta^\top X > u | \tilde{T} < \infty, T \leq t) \quad \text{and} \quad \mathcal{F}_t(\alpha, \beta, u) = P(\beta^\top X > u | \tilde{T} < \infty, T > t). \quad (31)$$

Similar to the discussion on (23), one can further express (31) as

$$\mathcal{T}_t(\alpha, \beta, u) = \frac{E[\pi(X) \{1 - S_u(t|X)\} \mathbb{I}(\beta^\top X > u)]}{E[\pi(X) \{1 - S_u(t|X)\}]} \quad (32)$$

and

$$\mathcal{F}_t(\alpha, \beta, u) = \frac{E\{\pi(X) S_u(t|X) \mathbb{I}(\beta^\top X > u)\}}{E\{\pi(X) S_u(t|X)\}}. \quad (33)$$

Thus, the time-dependent AUC is defined as

$$\mathcal{A}_t(\alpha, \beta) = \frac{E[\pi(X_1) \{1 - S_u(t|X_1)\} \pi(X_2) S_u(t|X_2) \mathbb{I}(\beta^\top X_1 > \beta^\top X_2)]}{E[\pi(X) \{1 - S_u(t|X)\}] E\{\pi(X) S_u(t|X)\}}. \quad (34)$$

Moreover, following a similar discussion to that in Section 3.3.1, the corrected time-dependent TPR, FPR, and AUC are given by (32), (33), and (34) with  $X$  being replaced by  $\tilde{X}$ , respectively. With the consistent estimators  $(\hat{\alpha}, \hat{\beta}, \hat{H}(\cdot))$ , the corresponding estimators are determined by

$$\hat{\mathcal{T}}_t(\hat{\alpha}, \hat{\beta}, u) = \frac{\frac{1}{n} \sum_{i=1}^n [\hat{\pi}(\tilde{X}_i) \{1 - \hat{S}_u(t|\tilde{X}_i)\} \mathbb{I}(\hat{\beta}^\top \tilde{X}_i > u)]}{\frac{1}{n} \sum_{i=1}^n [\hat{\pi}(\tilde{X}_i) \{1 - \hat{S}_u(t|\tilde{X}_i)\}]}$$

$$\hat{\mathcal{F}}_t(\hat{\alpha}, \hat{\beta}, u) = \frac{\frac{1}{n} \sum_{i=1}^n \left\{ \hat{\pi}(\tilde{X}_i) \hat{S}_u(t|\tilde{X}_i) \mathbb{I}(\hat{\beta}^\top \tilde{X}_i > u) \right\}}{\frac{1}{n} \sum_{i=1}^n \left\{ \hat{\pi}(\tilde{X}_i) \hat{S}_u(t|\tilde{X}_i) \right\}}$$

and

$$\hat{\mathcal{A}}_t(\hat{\alpha}, \hat{\beta}) = \frac{\frac{1}{n(n-1)} \sum_{i \neq j} [\hat{\pi}(\tilde{X}_i) \{1 - \hat{S}_u(t|\tilde{X}_i)\} \hat{\pi}(\tilde{X}_j) \hat{S}_u(t|\tilde{X}_j) \mathbb{I}(\hat{\beta}^\top \tilde{X}_i > \hat{\beta}^\top \tilde{X}_j)]}{\left[ \frac{1}{n} \sum_{i=1}^n \hat{\pi}(\tilde{X}_i) \{1 - \hat{S}_u(t|\tilde{X}_i)\} \right] \left[ \frac{1}{n} \sum_{i=1}^n \hat{\pi}(\tilde{X}_i) \hat{S}_u(t|\tilde{X}_i) \right]}.$$

Finally, at the end of this section, we establish the following theorems to justify the proposed estimators of time-independent/-dependent AUC, including consistency and asymptotic normality.

**Theorem 3.** Suppose that the conditions in Theorem 1 hold. With a fixed value  $u$ , as  $n \rightarrow \infty$ , the following applies:

(a) For the time-independent result in Section 3.3.1,

$$\hat{\mathcal{T}}^*(\hat{\alpha}, \hat{\beta}, u) \xrightarrow{p} \mathcal{T}(\alpha_0, \beta_0, u), \quad \hat{\mathcal{F}}^*(\hat{\alpha}, \hat{\beta}, u) \xrightarrow{p} \mathcal{F}(\alpha_0, \beta_0, u), \text{ and } \hat{\mathcal{A}}(\hat{\alpha}, \hat{\beta}) \xrightarrow{p} \mathcal{A}(\alpha_0, \beta_0);$$

(b) For the time-dependent result in Section 3.3.2,

$$\hat{\mathcal{T}}_t(\hat{\alpha}, \hat{\beta}, u) \xrightarrow{p} \mathcal{T}_t(\alpha_0, \beta_0, u), \quad \hat{\mathcal{F}}_t(\hat{\alpha}, \hat{\beta}, u) \xrightarrow{p} \mathcal{F}_t(\alpha_0, \beta_0, u),$$

and

$$\hat{\mathcal{A}}_t(\hat{\alpha}, \hat{\beta}) \xrightarrow{p} \mathcal{A}_t(\alpha_0, \beta_0).$$

**Theorem 4.** Suppose that conditions in Theorems 1 and 2 hold. Then as  $n \rightarrow \infty$ , the following applies:

(a) For the time-independent result in Section 3.3.1,

$$\sqrt{n} \{ \hat{\mathcal{A}}(\hat{\alpha}) - \mathcal{A}(\alpha_0) \} \xrightarrow{d} N(0, \sigma_{\pi}),$$

where  $\sigma_{\pi}$  is the asymptotic variance whose exact form is given in Appendix A.5;

(b) For the time-dependent result in Section 3.3.2,

$$\sqrt{n} \{ \hat{\mathcal{A}}_t(\hat{\alpha}, \hat{\beta}) - \mathcal{A}_t(\alpha_0, \beta_0) \} \xrightarrow{d} N(0, \sigma_{\text{TD}}),$$

where  $\sigma_{\text{TD}}$  is the asymptotic variance whose exact form is given in Appendix A.5.

## 4. Numerical Studies

Let  $n = 250$  or  $500$  denote the sample size. For a subject  $i = 1, \dots, n$ , let  $X_i$  be generated by the standard normal distribution. Let  $\alpha = 0.5$ ,  $\rho = 2$ , and  $\beta = 1$  be the true values of the parameters in (1) and (2), respectively. We consider the function  $H(t) = t$ . We then generate  $A_i$  and  $T_i$  by (1) and (2), respectively. We further independently generate the censoring time  $C_i$  from the uniform distribution in an interval  $[0, \tau_0]$ , where  $\tau_0$  is a pre-specified constant, such that the censoring rate is approximately 30%. In addition, we further apply measurement error model (3) to generate  $X_i^*$  for  $i = 1, \dots, n$ , where  $\epsilon_i$  is independently generated by the normal distribution with mean zero and variance  $\sigma_{\epsilon}^2 = 0.15, 0.35$ , and  $0.55$ , which reflect minor, moderate, or severe measurement error effects, respectively. Consequently, the collected dataset is denoted by  $\{ \{Y_i, \delta_i, X_i^*\} : i = 1, \dots, n \}$ . For each setting, we run the simulation 500 times repeatedly.

There are two goals in this study. First, we aim to estimate two parameters  $\alpha$  and  $\beta$  in (1) and (2), respectively. Based on two estimates  $\hat{\alpha}$  and  $\hat{\beta}$ , the next goal is to estimate the time-independent AUC (26) and the time-dependent AUC (34) at the time points  $t = 5$  and  $10$ . In addition to implementing the proposed method with measurement error correction, to see the

impact of the measurement error on estimation, we also examine the naive method, which is given by the same estimation procedure in Section 3 but with error-prone covariates  $X_i^*$  equipped.

The simulation results are summarized in Table 1, where we report the bias, the standard error (S.E.) obtained by repeated simulations, and the mean squared error (MSE). In general, we find that biases are increasing when  $\sigma_\epsilon^2$  becomes large, which implies that the measurement error may affect the estimation even though the correction is taken into account. In particular, larger biases of the estimated AUC indicate that the shape of the ROC curves slightly moves to the 45-degree straight line. For the comparisons of two estimation methods, the proposed method generally outperforms the naive method due to the smaller biases of the estimators, regardless of the change of the values of  $\sigma_\epsilon^2$  and sample sizes  $n$ . This indicates that the proposed method is valid to produce precise estimators. Note that the S.E. of the naive method is relatively smaller than the proposed method, as discussed in [22]; it is due to removing biases in point estimators and is a typical phenomenon in the framework of measurement error analysis. By the trade-off between bias and variation, we observe from the MSE that the proposed method is still better than the naive method with smaller MSE. In summary, the numerical results verify the validity of the proposed method and the measurement error correction.

**Table 1.** Simulation results. “Parameter” indicates the some parameters of interest; “Method” refers to the proposed and naive methods; “Bias” is the bias of the estimators; “S.E.” is the standard error; and “MSE” is the mean squared error.

| $\sigma_\epsilon^2$ | Parameter                         | Methods  | $n = 250$ |       |       | $n = 500$ |       |       |
|---------------------|-----------------------------------|----------|-----------|-------|-------|-----------|-------|-------|
| 0.15                | $\alpha$                          | Naive    | Bias      | S.E.  | MSE   | Bias      | S.E.  | MSE   |
|                     |                                   | Proposed | 0.163     | 0.308 | 0.121 | 0.159     | 0.230 | 0.078 |
|                     | $\beta$                           | Naive    | 0.012     | 0.325 | 0.106 | 0.009     | 0.238 | 0.057 |
|                     |                                   | Proposed | 0.170     | 0.231 | 0.082 | 0.165     | 0.208 | 0.070 |
|                     | $\mathcal{A}(\alpha, \beta)$      | Naive    | 0.018     | 0.266 | 0.071 | 0.013     | 0.220 | 0.049 |
|                     |                                   | Proposed | 0.108     | 0.033 | 0.013 | 0.096     | 0.030 | 0.010 |
|                     | $\mathcal{A}_5(\alpha, \beta)$    | Naive    | 0.014     | 0.056 | 0.003 | 0.007     | 0.047 | 0.002 |
|                     |                                   | Proposed | 0.115     | 0.051 | 0.016 | 0.108     | 0.047 | 0.014 |
|                     | $\mathcal{A}_{10}(\alpha, \beta)$ | Naive    | 0.016     | 0.060 | 0.004 | 0.013     | 0.057 | 0.003 |
|                     |                                   | Proposed | 0.114     | 0.049 | 0.015 | 0.103     | 0.041 | 0.012 |
|                     |                                   | Naive    | 0.013     | 0.062 | 0.004 | 0.009     | 0.056 | 0.003 |
|                     |                                   | Proposed |           |       |       |           |       |       |
| 0.35                | $\alpha$                          | Naive    | 0.195     | 0.321 | 0.141 | 0.178     | 0.255 | 0.097 |
|                     |                                   | Proposed | 0.020     | 0.345 | 0.119 | 0.017     | 0.269 | 0.073 |
|                     | $\beta$                           | Naive    | 0.195     | 0.264 | 0.108 | 0.181     | 0.229 | 0.085 |
|                     |                                   | Proposed | 0.023     | 0.281 | 0.080 | 0.020     | 0.245 | 0.060 |
|                     | $\mathcal{A}(\alpha, \beta)$      | Naive    | 0.125     | 0.047 | 0.018 | 0.114     | 0.042 | 0.015 |
|                     |                                   | Proposed | 0.017     | 0.060 | 0.004 | 0.014     | 0.055 | 0.003 |
|                     | $\mathcal{A}_5(\alpha, \beta)$    | Naive    | 0.123     | 0.060 | 0.019 | 0.112     | 0.055 | 0.016 |
|                     |                                   | Proposed | 0.020     | 0.071 | 0.005 | 0.017     | 0.066 | 0.005 |
|                     | $\mathcal{A}_{10}(\alpha, \beta)$ | Naive    | 0.121     | 0.054 | 0.018 | 0.116     | 0.050 | 0.016 |
|                     |                                   | Proposed | 0.019     | 0.069 | 0.005 | 0.016     | 0.063 | 0.004 |
|                     |                                   | Naive    |           |       |       |           |       |       |
|                     |                                   | Proposed |           |       |       |           |       |       |
| 0.55                | $\alpha$                          | Naive    | 0.214     | 0.348 | 0.167 | 0.193     | 0.285 | 0.118 |
|                     |                                   | Proposed | 0.027     | 0.362 | 0.132 | 0.023     | 0.307 | 0.095 |
|                     | $\beta$                           | Naive    | 0.226     | 0.287 | 0.133 | 0.201     | 0.266 | 0.111 |
|                     |                                   | Proposed | 0.028     | 0.301 | 0.091 | 0.025     | 0.278 | 0.078 |
|                     | $\mathcal{A}(\alpha, \beta)$      | Naive    | 0.133     | 0.066 | 0.022 | 0.128     | 0.059 | 0.020 |
|                     |                                   | Proposed | 0.021     | 0.078 | 0.006 | 0.019     | 0.069 | 0.005 |
|                     | $\mathcal{A}_5(\alpha, \beta)$    | Naive    | 0.136     | 0.068 | 0.023 | 0.124     | 0.063 | 0.019 |
|                     |                                   | Proposed | 0.026     | 0.079 | 0.007 | 0.024     | 0.074 | 0.006 |
|                     | $\mathcal{A}_{10}(\alpha, \beta)$ | Naive    | 0.134     | 0.060 | 0.022 | 0.128     | 0.057 | 0.020 |
|                     |                                   | Proposed | 0.025     | 0.066 | 0.005 | 0.021     | 0.069 | 0.005 |
|                     |                                   | Naive    |           |       |       |           |       |       |
|                     |                                   | Proposed |           |       |       |           |       |       |

## 5. Summary

In this paper, we study the transformation model to analyze the survival data with the cure fraction and its extension to the estimation of the ROC curve. In addition, we also explore the measurement error in covariates or biomarkers, which is one type of noisy

data. To deal with measurement error effects and derive a reliable estimator, we propose a corrected likelihood function whose expectation recovers to the likelihood function under the true covariates. In addition, we also propose a measurement error correction strategy to estimate the time-independent/time-dependent AUC. To derive the optimization of (19), we introduce the EM algorithm with finite stepwise iterations to obtain the estimators, but one should keep in mind that the computational implementation is not unique, and some alternative strategies can be adopted, such as quasi-Newton methods or damped Anderson acceleration, when the dimension of variables is large or the settings are complex. The other contribution of this paper is the focus on the establishment of the asymptotic properties for the proposed estimator with rigorous derivations. While we have no real data application in the current stage, it is also expected that the proposed method is valid to handle the relevant data, since simulation studies show that the performance of the proposed method is satisfactory.

The current development focuses on the estimation of the parameters, as well as the AUC, and the handling of measurement error problems, but some further issues related to the AUC could be explored in depth. For example, the current setting involves the cure fraction, which may induce long follow-up periods for the survival time. An interesting but unsolved concern is to understand whether the dynamic AUC estimation is stabilized and the corresponding impact caused by measurement error. Exploring this issue in depth might make the diagnosis of the disease more precise. In the next stage of our research, we wish to explore this issue from a theoretical perspective and provide rigorous justification. In addition to the time-dependent AUC, as mentioned by a referee, the other challenging feature in applications is the time-varying covariates. While this issue was studied in the literature (e.g., [28]), the corresponding applications to AUC estimation seem to be rarely discussed. Moreover, in the presence of measurement error in covariates, the other challenge includes the characterization of time-dependent covariates and the corresponding measurement error correction. We expect that the current method can be extended to handle this complex structure in the near future.

**Funding:** This research was funded by National Science and Technology Council, Taiwan, with grant number 112-2118-M-004-005-MY2.

**Data Availability Statement:** All the data in this paper are simulated data generated by the procedure in Section 4.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A. Theoretical Justification

### Appendix A.1. Regularity Conditions

- (C1)  $\Theta$  is a compact set, and the true parameter value  $(\alpha_0, \beta_0)$  is an interior point of  $\Theta$ .
- (C2) Let  $\tau$  be the finite maximum support of the failure time.
- (C3)  $\{A_i, Y_i, X_i\}$  are independent and identically distributed for  $i = 1, \dots, n$ .
- (C4) The covariates or biomarkers  $X_i$  are bounded.
- (C5) Censoring time  $C_i$  is noninformative. That is, the failure time  $T_i$  and the censoring time  $C_i$  are independent, given the covariate  $X_i$ .

Condition (C1) is a basic condition that is used to derive the maximizer of the target function. (C2) to (C5) are standard conditions for survival analysis, which allow us to obtain the sum of i.i.d. random variables and hence to derive the asymptotic properties of the estimators.

### Appendix A.2. Proof of Theorem 1

**Proof.** Note that the function  $H(\cdot)$  is parameterized as  $H^*(t) \triangleq \frac{1}{n} \sum_{j=1}^n \mathbb{I}(Y_j \leq t) \exp(\gamma_j)$ , and the corresponding estimator is given by  $\hat{H}(t) = \frac{1}{n} \sum_{j=1}^n \mathbb{I}(Y_j \leq t) \exp(\hat{\gamma}_j)$ . Following Helly's Lemma (e.g., [29], Section 2.1) and a similar discussion in [30], we can show that  $H^*(t)$  converges to  $H_0(t)$ , provided that  $\gamma_j$  is replaced by the true value  $\gamma_{j0}$  for  $j = 1, \dots, p$ . As a result, to show the consistency of  $\hat{H}(t)$  and  $H_0(t)$ , it suffices to examine the relationship between  $\hat{H}(t)$  and  $H^*(t)$ ; thus, we can explore the asymptotic behavior of  $\hat{\gamma}_j$ .

In the following derivation, we aim to prove that

$$\left\| \left( \hat{\alpha}^\top, \hat{\beta}^\top, \hat{\gamma}^\top \right)^\top - \left( \alpha_0^\top, \beta_0^\top, \gamma_0^\top \right)^\top \right\|_2 = O_p \left( \frac{1}{\sqrt{n}} \right). \quad (\text{A1})$$

Specifically, if (A1) holds, then we conclude that  $\|\hat{\alpha} - \alpha_0\|_2 \xrightarrow{p} 0$ ,  $\|\hat{\beta} - \beta_0\|_2 \xrightarrow{p} 0$  and  $\|\hat{H}(t) - H^*(t)\|_\infty \xrightarrow{p} 0$  as  $n \rightarrow \infty$ .

To prove (A1), we follow the strategy in [24] and start by showing that

$$P \left\{ \sup_{\|u\|_2 = \mathcal{B}} \hat{\ell}^* \left( \alpha_0 + \frac{u_1}{\sqrt{n}}, \beta_0 + \frac{u_2}{\sqrt{n}}, \gamma_0 + \frac{u_3}{\sqrt{n}} \right) < \hat{\ell}^*(\alpha_0, \beta_0, \gamma_0) \right\} > 1 - \epsilon \quad (\text{A2})$$

for every  $\epsilon > 0$  and a given value  $\mathcal{B}$ , where  $U = (u_1^\top, u_2^\top, u_3^\top)^\top$ , with  $u_1 = \sqrt{n}(\hat{\alpha} - \alpha_0)$ ,  $u_2 = \sqrt{n}(\hat{\beta} - \beta_0)$ , and  $u_3 = \sqrt{n}(\hat{\gamma} - \gamma_0)$ .

We write  $\Psi(u_1, u_2, u_3) = \hat{\ell}^* \left( \alpha_0 + \frac{u_1}{\sqrt{n}}, \beta_0 + \frac{u_2}{\sqrt{n}}, \gamma_0 + \frac{u_3}{\sqrt{n}} \right) - \hat{\ell}^*(\alpha_0, \beta_0, \gamma_0)$ . By the Taylor series expansion around  $u_1 = u_2 = u_3 = 0$ , we have

$$\begin{aligned} \Psi(u_1, u_2, u_3) &= \frac{u_1^\top}{\sqrt{n}} S_\alpha(\alpha_0, \beta_0, \gamma_0) + \frac{u_2^\top}{\sqrt{n}} S_\beta(\alpha_0, \beta_0, \gamma_0) + \frac{u_3^\top}{\sqrt{n}} S_\gamma(\alpha_0, \beta_0, \gamma_0) \\ &\quad + \frac{1}{n} u_1^\top I_\alpha(\alpha_0, \beta_0, \gamma_0) u_1 + \frac{1}{n} u_2^\top I_\beta(\alpha_0, \beta_0, \gamma_0) u_2 + \frac{1}{n} u_3^\top I_\gamma(\alpha_0, \beta_0, \gamma_0) u_3 \\ &\quad + \frac{1}{n} u_2^\top I_{\beta\gamma}(\alpha_0, \beta_0, \gamma_0) u_3 + o_p(1), \end{aligned} \quad (\text{A3})$$

where  $S_\alpha(\cdot)$ ,  $S_\beta(\cdot)$ , and  $S_\gamma(\cdot)$  are the first-order derivatives of  $\Psi(\cdot)$  with respect to  $u_1$ ,  $u_2$ , and  $u_3$ , respectively;  $I_\alpha(\cdot)$ ,  $I_\beta(\cdot)$ , and  $I_\gamma(\cdot)$  are the second-order derivatives of  $\Psi(\cdot)$  with respect to  $u_1$ ,  $u_2$ , and  $u_3$ , respectively; and  $I_{\beta\gamma}$  is the derivative of  $\Psi(\cdot)$  with respect to  $u_2$  and  $u_3$ .

We now separately examine each term in (A3). Note that  $(\alpha_0, \beta_0, \gamma_0)$  is the maximizer of  $E(\ell)$ , and  $E(\ell) = E(\hat{\ell}^*)$  according to our development. This implies that  $(\alpha_0, \beta_0, \gamma_0)$  is also the maximizer of  $E(\hat{\ell}^*)$  or, equivalently, the solutions of  $E\{S_\alpha(\alpha_0, \beta_0, \gamma_0)\} = 0$ ,  $E\{S_\beta(\alpha_0, \beta_0, \gamma_0)\} = 0$ , and  $E\{S_\gamma(\alpha_0, \beta_0, \gamma_0)\} = 0$ . Therefore, we have

$$\begin{aligned} \frac{u_1^\top}{\sqrt{n}} S_\alpha(\alpha_0, \beta_0, \gamma_0) &= u_1^\top O_p(1), \\ \frac{u_2^\top}{\sqrt{n}} S_\beta(\alpha_0, \beta_0, \gamma_0) &= u_2^\top O_p(1), \text{ and} \\ \frac{u_3^\top}{\sqrt{n}} S_\gamma(\alpha_0, \beta_0, \gamma_0) &= u_3^\top O_p(1). \end{aligned}$$

For the second-order derivative  $I_\alpha \triangleq \frac{\partial^2 \hat{\ell}^*}{\partial \alpha \partial \alpha^\top}$ , its explicit form is given by

$$\begin{aligned} I_\alpha &= - \left( \sum_{i=1}^n \left[ \frac{X_i^* X_i^{*\top} \exp(\alpha^\top X_i^*) \{1 + \exp(\alpha^\top X_i^*)\} - X_i^* X_i^{*\top} \exp(\alpha^\top X_i^*)}{\{1 + \exp(\alpha^\top X_i^*)\}^2} \right] + \Sigma_\epsilon \right) \\ &= \frac{\partial}{\partial \alpha^\top} \left( \frac{\partial \hat{\ell}^*}{\partial \alpha} \right). \end{aligned} \quad (\text{A4})$$

By the Law of Large Numbers, we have that  $\frac{1}{n} I_\alpha \xrightarrow{p} \mathcal{I}_\alpha = E \left( \frac{1}{n} I_\alpha \right) = E \left\{ E \left( \frac{1}{n} I_\alpha \mid X_i \right) \right\}$  as  $n \rightarrow \infty$ . As a result, to find  $E(I_\alpha)$ , it suffices to compute  $E(I_\alpha \mid X_i) = E \left\{ \frac{\partial}{\partial \alpha^\top} \left( \frac{\partial \hat{\ell}^*}{\partial \alpha} \right) \mid X_i \right\} = \frac{\partial}{\partial \alpha^\top} E \left( \frac{\partial \hat{\ell}^*}{\partial \alpha} \mid X_i \right)$ , where the last equality is based on the interchange of the derivative and integral.

Taking the derivative of (19) with respect to  $\alpha$  and then taking the conditional expectation give

$$E \left( \frac{\partial \hat{\ell}^*}{\partial \alpha} \mid X_i \right) = \sum_{i=1}^n \left\{ \xi_i X_i - \frac{\exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha) \times (X_i + \Sigma_\epsilon \alpha)}{1 + \exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha)} \right\} + \Sigma_\epsilon \alpha,$$

which also implies that

$$\begin{aligned} E \left( \frac{\partial^2 \hat{\ell}^*}{\partial \alpha \partial \alpha^\top} \mid X_i \right) &= \frac{\partial}{\partial \alpha^\top} E \left( \frac{\partial \hat{\ell}^*}{\partial \alpha} \mid X_i \right) \\ &= - \sum_{i=1}^n \frac{\exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha) \{ \|X_i + \Sigma_\epsilon \alpha\|_2^2 - \Sigma_\epsilon \} - \Sigma_\epsilon}{\{1 + \exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha)\}^2}. \end{aligned}$$

As a result, by the Law of Large Numbers and Condition (C3),  $\mathcal{I}_\alpha$  is determined by

$$\mathcal{I}_\alpha = E \left[ \frac{- \exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha) \{ \|X_i + \Sigma_\epsilon \alpha\|_2^2 - \Sigma_\epsilon \} + \Sigma_\epsilon}{\{1 + \exp(\alpha^\top X_i + \frac{1}{2} \alpha^\top \Sigma_\epsilon \alpha)\}^2} \right].$$

By Conditions (C1) and (C4), the covariates  $X_i$  and the parameter  $\alpha$  are bounded, and  $\Sigma_\epsilon$  is positive definite, which indicates that  $\mathcal{I}_\alpha$  is bounded componentwise and its negative definite ensures that the maximizer of  $\alpha$  exists.

Next, we explore  $I_\beta$ . By taking the derivative of  $\hat{\ell}^*$  with respect to  $\beta$ , we have

$$\begin{aligned} \frac{\partial \hat{\ell}^*}{\partial \beta} &= \sum_{i=1}^n \left[ \delta_i \frac{X_i^* \exp(-\beta^\top X_i^*)}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp(-\beta^\top X_i^*)} + \Sigma_\epsilon \beta - \frac{A_i}{\rho} X_i^* \right. \\ &\quad \left. + \frac{A_i}{\rho} \frac{X_i^* \exp(-\beta^\top X_i^*)}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp(-\beta^\top X_i^*)} + \frac{A_i}{\rho} \Sigma_\epsilon \beta \right]. \end{aligned}$$

Note that the moment-generating function gives

$$E \left\{ \exp(-\beta^\top X_i^*) \mid X_i \right\} = \exp \left( -\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta \right) \quad (\text{A5})$$

and

$$E \left\{ -X_i^* \exp(\beta^\top X_i^*) \mid X_i \right\} = \exp \left( -\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta \right) \times (-X_i + \Sigma_\epsilon \beta),$$

then, the conditional expectation of  $\frac{\partial \hat{\ell}^*}{\partial \beta}$  given  $X_i$  is

$$E\left(\frac{\partial \hat{\ell}^*}{\partial \beta} \middle| X_i\right) = \sum_{i=1}^n \left[ \delta_i \frac{(X_i - \Sigma_\epsilon \beta) \exp\left(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta\right)}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp\left(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta\right)} + \Sigma_\epsilon \beta - \frac{A_i}{\rho} X_i \right. \\ \left. + \frac{A_i}{\rho} \frac{(X_i - \Sigma_\epsilon \beta) \exp\left(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta\right)}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp\left(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta\right)} + \frac{A_i}{\rho} \Sigma_\epsilon \beta \right].$$

This further shows that

$$E\left(\frac{\partial^2 \hat{\ell}^*}{\partial \beta \partial \beta^\top} \middle| X_i\right) = \sum_{i=1}^n \left\{ \left( \delta_i + \frac{A_i}{\rho} \right) D_1 + \frac{A_i}{\rho} \Sigma_\epsilon \right\} \text{ and } E\left(\frac{\partial^2 \hat{\ell}^*}{\partial \beta \partial \gamma_j^\top} \middle| X_i\right) = \sum_{i=1}^n \left\{ \left( \delta_i + \frac{A_i}{\rho} \right) D_2 \right\},$$

where

$$D_1 = \frac{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) (\| -X_i + \Sigma_\epsilon \beta \|_2^2 + \Sigma_\epsilon) \exp\left(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta\right)}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp\left(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta\right)} \\ + \frac{\exp(-2\beta^\top X_i + \beta^\top \Sigma_\epsilon \beta) \Sigma_\epsilon}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp\left(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta\right)}$$

and

$$D_2 = \frac{\rho \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) (-X_i + \Sigma_\epsilon \beta) \exp\left(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta\right)}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp\left(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta\right)}.$$

Consequently, by the Law of Large Numbers, we have that

$$\mathcal{I}_\beta = E\left(\frac{1}{n} \frac{\partial^2 \hat{\ell}^*}{\partial \beta \partial \beta^\top} \middle| X_i\right) = E\left\{ \left( \delta_i + \frac{A_i}{\rho} \right) D_1 + \frac{A_i}{\rho} \Sigma_\epsilon \right\}$$

and  $\mathcal{I}_{\beta\gamma}$  is the matrix of  $E\left(\frac{\partial^2 \hat{\ell}^*}{\partial \beta \partial \gamma_j^\top} \middle| X_i\right)$  with stack in row for  $j = 1, \dots, n$ , where

$$E\left(\frac{\partial^2 \hat{\ell}^*}{\partial \beta \partial \gamma_j^\top} \middle| X_i\right) = E\left\{ \left( \delta_i + \frac{A_i}{\rho} \right) D_2 \right\},$$

and they are bounded elementally due to positive definite matrix  $\Sigma_\epsilon$  and Conditions (C1) and (C4).

Finally, we examine  $\mathcal{I}_\gamma$ . Taking the first derivative of  $\hat{\ell}^*$  with respect to  $\gamma_j$  yields

$$\frac{\partial \hat{\ell}^*}{\partial \gamma_j} = \sum_{i=1}^n \left\{ \delta_i - \frac{\rho \mathbb{I}(Y_i < Y_j) \exp(\gamma_j)}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp(-\beta^\top X_i^*)} \times \left( \delta_i + \frac{\xi_i^*}{\rho} \right) \right\}.$$

Additionally, by (A5), we have

$$E\left\{\frac{\partial \widehat{\ell}^*}{\partial \gamma_j} \middle| X_i\right\} = \sum_{i=1}^n \left\{ \delta_i - \frac{\rho \mathbb{I}(Y_i < Y_j) \exp(\gamma_j)}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta)} \times \left( \delta_i + \frac{A_i}{\rho} \right) \right\}.$$

This also implies that

$$E\left\{\frac{\partial^2 \widehat{\ell}^*}{\partial \gamma_j^2} \middle| X_i\right\} = - \sum_{i=1}^n \left\{ \frac{\rho^2 \mathbb{I}(Y_i < Y_j) \exp(2\gamma_j)}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta)} \times \left( \delta_i + \frac{A_i}{\rho} \right) \right\}$$

and

$$E\left\{\frac{\partial^2 \widehat{\ell}^*}{\partial \gamma_j \partial \gamma_l} \middle| X_i\right\} = - \sum_{i=1}^n \left\{ \frac{\rho^2 \mathbb{I}(Y_i < Y_j) \mathbb{I}(Y_i < Y_l) \exp(\gamma_l + \gamma_j)}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta)} \times \left( \delta_i + \frac{A_i}{\rho} \right) \right\}$$

for  $j \neq l$ . Therefore, by the Law of Large Numbers, we have  $\mathcal{I}_\gamma$  with diagonal entries

$$E\left\{\frac{\partial^2 \widehat{\ell}^*}{\partial \gamma_j^2}\right\} = E\left\{- \frac{\rho^2 \mathbb{I}(Y_i < Y_j) \exp(2\gamma_j)}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta)} \times \left( \delta_i + \frac{A_i}{\rho} \right) \right\}$$

and off-diagonal elements

$$E\left\{\frac{\partial^2 \widehat{\ell}^*}{\partial \gamma_j \partial \gamma_l}\right\} = E\left\{- \frac{\rho^2 \mathbb{I}(Y_i < Y_j) \mathbb{I}(Y_i < Y_l) \exp(\gamma_l + \gamma_j)}{\rho \sum_{j=1}^n \mathbb{I}(Y_i < Y_j) \exp(\gamma_j) + \exp(-\beta^\top X_i + \frac{1}{2} \beta^\top \Sigma_\epsilon \beta)} \times \left( \delta_i + \frac{A_i}{\rho} \right) \right\}$$

for  $j \neq l$ .

Finally, when  $n$  goes to infinity, the quadratic terms in (A3) converge in probability to

$$\left(u_1^\top, u_2^\top, u_3^\top\right)^\top \begin{pmatrix} \mathcal{I}_\alpha & 0 & 0 \\ 0 & \mathcal{I}_\beta & \mathcal{I}_{\beta\gamma} \\ 0 & \mathcal{I}_{\beta\gamma} & \mathcal{I}_\gamma \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} =: U^\top \mathcal{I} U,$$

which is equal to  $O(\mathcal{B}^2)$  because  $\mathcal{I}$  is bounded and negative definite. Since  $\mathcal{B}^2$  dominates  $\mathcal{B}$  in the linear terms of (A3), for sufficiently large  $\mathcal{B}$  and  $\|U\|_2 = \mathcal{B}$ , we have

$$\widehat{\ell}^*\left(\alpha_0 + \frac{u_1}{\sqrt{n}}, \beta_0 + \frac{u_2}{\sqrt{n}}, \gamma_0 + \frac{u_3}{\sqrt{n}}\right) - \widehat{\ell}^*(\alpha_0, \beta_0, \gamma_0) < 0.$$

Therefore, (A2) holds; thus, (A1) is verified.  $\square$

### Appendix A.3. Proof of Theorem 2

**Proof.** In this proof, we adopt Theorem 3.3.1 in [31] to derive the asymptotic distribution. A similar discussion was also presented by [32].

Let  $\theta_0 \triangleq (\alpha_0, \beta_0, H_0)$  and  $\widehat{\theta} \triangleq (\widehat{\alpha}, \widehat{\beta}, \widehat{H})$ . Moreover, define  $S = \frac{\partial \widehat{\ell}^*}{\partial \theta}$  and  $\mathcal{S} = E(S)$ . According to Theorem 3.3.1 in [31], we need to check the following:

- (a)  $\sqrt{n}(S - \mathcal{S})(\theta_0) \xrightarrow{d} Z$ , where  $Z$  is a tight random element;
- (b) The map  $\theta \rightarrow \mathcal{S}(\theta)$  is Fréchet differentiable at  $\theta_0$  with a continuously invertible derivative  $\nabla \mathcal{S}(\theta_0)$ , where  $\nabla f(x)$  denotes an operator of the derivative of  $f$  with respect to  $x$ ;
- (c)  $\mathcal{S}(\theta_0) = 0$ , and  $\hat{\theta}$  satisfies  $\mathcal{S}(\hat{\theta}) = o_p\left(\frac{1}{\sqrt{n}}\right)$ .

If Conditions (a)–(c) are verified, then we conclude that

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} -\{\nabla \mathcal{S}(\theta_0)\}^{-1} Z \text{ as } n \rightarrow \infty.$$

Lastly, we examine Conditions (a)–(c) separately.

**Check Condition (a):**

Since  $\theta_0$  is the true value, we automatically have  $\mathcal{S}(\theta_0) = 0$ . Then,  $(S - \mathcal{S})(\theta_0) = \frac{1}{n} \sum_{i=1}^n (D_{i,1} + D_{i,2} + D_{i,3})$ , where

$$D_{i,1} = X_i^* - \frac{X_i^* \exp(\alpha_0^\top X_i^*)}{1 + \exp(\alpha_0^\top X_i^*)} + \Sigma_\epsilon \alpha_0,$$

$$D_{i,2} = \left( \frac{X_i^* \exp(-\beta_0^\top X_i^*)}{\rho H_0(Y_i) + \exp(-\beta_0^\top X_i^*)} + \Sigma_\epsilon \beta_0 \right) \left( \frac{A_i}{\rho} + \delta_i \right),$$

and

$$D_{i,3} = \delta_i dH_0(Y_i) - \frac{\rho dH_0(Y_i)}{\rho H_0(Y_i) + \exp(-\beta_0^\top X_i^*)} \left( \frac{A_i}{\rho} + \delta_i \right).$$

By Conditions (C1) and (C4),  $D_{i,1}$  is Donsker bounded because it belongs to the finite-dimensional class of a measurable score function. In addition,  $D_{i,2}$  and  $D_{i,3}$  are bounded real valued functions on the bounded support due to Conditions (C2) and (C4). This indicates that  $D_{i,2}$  and  $D_{i,3}$  are Donsker (e.g., [29], p. 270). Therefore, by Example 2.10.7 in [31], we conclude that  $(S - \mathcal{S})(\theta_0)$  is also Donsker. Consequently, by Section 2.8.2 in [31] and by Theorem 19.3 in [29], we obtain that as  $n \rightarrow \infty$ ,  $\sqrt{n}(S - \mathcal{S})(\theta_0)$  converges in distribution to a tight random element, denoted by  $Z$ .

**Check Condition (b):**

The first part is to show Fréchet differentiability. That is, it suffices to prove that (e.g., [29], p. 297)

$$\left\| \mathcal{S}(\theta_0 + v) - \mathcal{S}(\theta_0) - \nabla \mathcal{S}(v) \right\| = o(\|v\|) \text{ as } \|v\| \rightarrow 0. \quad (\text{A6})$$

Let  $\theta \triangleq (\alpha, \beta, H)$ , and let  $\theta_0$  denote the true value of  $\theta$ . We write  $\theta_r = (\alpha + v_1, \beta + v_2, H_r(v_3))$ , with  $H_r(v_3)(t) = \int_0^t (1 + rv_3) dH(u)$ , where  $v_1, v_2 \in \mathbb{R}^p$  and  $v_3$  satisfies bound variation on  $[0, \tau]$ . Define  $v \equiv (v_1, v_2, v_3)$ .

On one hand, by the Taylor series expansion on  $\mathcal{S}(\theta_0 + rv)$  around  $r = 0$ , we have

$$\begin{aligned}
& \mathcal{S}(\theta_0 + r\nu) - \mathcal{S}(\theta_0) \\
&= \nu_1 E \left[ X_i^* - \frac{X_i^* \exp\{(\alpha_0 + r^* \nu_1)^\top X_i^*\}}{1 + \exp\{(\alpha_0 + r^* \nu_1)^\top X_i^*\}} + \Sigma_\epsilon(\alpha_0 + r\nu_1) \right] (r - 0) \\
&\quad + \nu_2 E \left[ \frac{X_i^* \exp\{-(\beta_0 + r^* \nu_2)^\top X_i^*\}}{\rho H_0(Y_i) + \exp\{-(\beta_0 + r^* \nu_2)^\top X_i^*\}} \cdot \left( \delta_i + \frac{A_i}{\rho} \right) \right. \\
&\quad \left. + \Sigma_\epsilon(\beta_0 + r^* \nu_2) \cdot \left( \delta_i + \frac{A_i}{\rho} \right) \right] (r - 0) \\
&\quad + \nu_3 E \left[ \delta_i \nu_3(Y_i) - \frac{\rho \int_0^{Y_i} \nu_3(u) dH_0(u)}{\rho H_{r^*}(Y_i) + \exp(-\beta_0^\top X_i^*)} \cdot \left( \delta_i + \frac{A_i}{\rho} \right) \right] (r - 0) + o(r) \\
&\triangleq r \cdot \{ \nu_1 \mathcal{S}_\alpha + \nu_2 \mathcal{S}_\beta + \mathcal{S}_H(\nu_3) \} + o(r).
\end{aligned} \tag{A7}$$

On the other hand, by taking the derivative of  $\mathcal{S}$  based on the definition in p. 296 of [29], we have

$$\begin{aligned}
\nabla \mathcal{S}(\nu) &= \nu_1 E \left[ X_i^* - \frac{X_i^* \exp\{\alpha_0^\top X_i^*\}}{1 + \exp\{\alpha_0^\top X_i^*\}} + \Sigma_\epsilon \alpha \right] \\
&\quad + \nu_2 E \left[ \frac{X_i^* \exp\{-\beta^\top X_i^*\}}{\rho H(Y_i) + \exp\{-\beta^\top X_i^*\}} \cdot \left( \delta_i + \frac{A_i}{\rho} \right) + \Sigma_\epsilon \beta \cdot \left( \delta_i + \frac{A_i}{\rho} \right) \right] \\
&\quad + \nu_3 E \left[ \delta_i \nu_3(Y_i) - \frac{\rho \int_0^{Y_i} \nu_3(u) dH(u)}{\rho H(Y_i) + \exp(-\beta^\top X_i^*)} \cdot \left( \delta_i + \frac{A_i}{\rho} \right) \right].
\end{aligned} \tag{A8}$$

Finally, by combining (A7) and (A8), we obtain (A6) as  $r \rightarrow 0$ , because of Theorem 1 and because  $r^* \rightarrow 0$  as  $r \rightarrow 0$ .

The second part is to discuss the continuous invertibility of  $\nabla \mathcal{S}(\theta_0)$ . Similar to the discussion in [32], the continuous invertibility of the Fréchet derivative can be justified by showing that there exists a constant  $\zeta$  such that

$$\inf_{\theta} \frac{\|\nabla \mathcal{S}(\theta_0)\|}{\|\theta\|} > \zeta. \tag{A9}$$

Since  $\nabla \mathcal{S}(\theta_0)$  can be expressed as a linear combination of the three  $\sigma$ -operators, where three terms in  $\nabla \mathcal{S}(\theta_0)$  are continuously differentiable functions that map to a finite-dimensional space. Therefore, we conclude that  $\nabla \mathcal{S}(\theta_0)$  is the summation of invertible and compact operators; thus, (A9) holds by the similar arguments in [32]. Consequently,  $\nabla \mathcal{S}(\theta_0)$  is verified to be continuously differentiable.

#### Check Condition (c):

The first claim,  $\mathcal{S}(\theta_0) = 0$ , has been verified by the argument of Condition (a). In addition,  $\hat{\theta}$  is the maximizer of  $\mathcal{S}(\theta)$ , and  $\hat{\theta}$  is the consistent estimator due to Theorem 1. Therefore, the second claim,  $\mathcal{S}(\hat{\theta}) = o_p(n^{-1/2})$ , holds.  $\square$

#### Appendix A.4. Proof of Theorem 3

**Proof.** In this appendix, we mainly prove results (a) and (b) separately.

Proof of part (a)

In this proof, we only show that, as  $n \rightarrow \infty$ ,

$$\hat{\mathcal{A}}(\hat{\alpha}, \hat{\beta}) \xrightarrow{p} \mathcal{A}(\alpha_0, \beta_0), \tag{A10}$$

and we omit the proofs of  $\hat{T}^*(\hat{\alpha}, \hat{\beta}, u)$  and  $\hat{F}^*(\hat{\alpha}, \hat{\beta}, u)$  because the derivations are similar to those for (A10).

We write  $\hat{\mathcal{A}}(\hat{\alpha}, \hat{\beta}) = \frac{\hat{\mathcal{A}}_n(\hat{\alpha}, \hat{\beta})}{\hat{\mathcal{A}}_d(\hat{\alpha}, \hat{\beta})}$ , where

$$\hat{\mathcal{A}}_d(\hat{\alpha}, \hat{\beta}) = \left\{ \frac{1}{n} \sum_{i=1}^n \hat{\pi}(\tilde{X}_i) \right\} \left[ \frac{1}{n} \sum_{i=1}^n \{1 - \hat{\pi}(\tilde{X}_i)\} \right]$$

and

$$\hat{\mathcal{A}}_n(\hat{\alpha}, \hat{\beta}) = \frac{1}{n(n-1)} \sum_{i \neq j} \left[ \hat{\pi}(\tilde{X}_i) \{1 - \hat{\pi}(\tilde{X}_j)\} \mathbb{I}(\hat{\beta}^\top \tilde{X}_i > \hat{\beta}^\top \tilde{X}_j) \right].$$

In the following, we examine  $\hat{\mathcal{A}}_d(\hat{\alpha}, \hat{\beta})$  and  $\hat{\mathcal{A}}_n(\hat{\alpha}, \hat{\beta})$  separately.

**Claim 1:** Let  $\mathcal{A}_d(\alpha_0, \beta_0) = E\{\pi(X)\}E\{1 - \pi(X)\}$ . Show that as  $n \rightarrow \infty$ ,

$$\hat{\mathcal{A}}_d(\hat{\alpha}, \hat{\beta}) \xrightarrow{p} \mathcal{A}_d(\alpha_0, \beta_0).$$

Note that by (28), we have  $E\{\pi(X)\} = E\{\pi(\tilde{X})\}$ , yielding that  $\mathcal{A}_d(\alpha_0, \beta_0) = E\{\pi(\tilde{X})\}E\{1 - \pi(\tilde{X})\}$ . By Theorem 1, which shows that  $\hat{\alpha}$  is the consistent estimator of  $\alpha_0$ , we have  $\hat{\pi}(X_i) \xrightarrow{p} \pi(X_i)$  as  $n \rightarrow \infty$ . On the other hand, by Condition (C3), which shows that  $X_i$  is independent, applying the Law of Large Number gives that as  $n \rightarrow \infty$ ,  $\frac{1}{n} \sum_{i=1}^n \hat{\pi}(\tilde{X}_i) \xrightarrow{p} E\{\pi(\tilde{X})\}$  and  $\frac{1}{n} \sum_{i=1}^n \{1 - \hat{\pi}(\tilde{X}_i)\} \xrightarrow{p} E\{1 - \pi(\tilde{X})\}$ . Thus, Claim 1 holds.

**Claim 2:** Let  $\mathcal{A}_n(\alpha_0, \beta_0) = E[\pi(X_1)\{1 - \pi(X_2)\}\mathbb{I}(\beta^\top X_1 > \beta^\top X_2)]$ . Show that as  $n \rightarrow \infty$ ,

$$\hat{\mathcal{A}}_n(\hat{\alpha}, \hat{\beta}) \xrightarrow{p} \mathcal{A}_n(\alpha_0, \beta_0).$$

Similar to Claim 1, we can obtain that  $\mathcal{A}_n(\alpha_0, \beta_0)$  is equal to  $E[\pi(\tilde{X}_1)\{1 - \pi(\tilde{X}_2)\}\mathbb{I}(\beta^\top \tilde{X}_1 > \beta^\top \tilde{X}_2)]$  due to (29). By adding and subtracting an additional term, we have

$$\begin{aligned} \hat{\mathcal{A}}_n(\hat{\alpha}, \hat{\beta}) - \mathcal{A}_n(\alpha_0, \beta_0) &= \left\{ \hat{\mathcal{A}}_n(\hat{\alpha}, \hat{\beta}) - \tilde{\mathcal{A}}_n(\alpha_0, \beta_0) \right\} + \left\{ \tilde{\mathcal{A}}_n(\alpha_0, \beta_0) - \mathcal{A}_n(\alpha_0, \beta_0) \right\} \\ &\triangleq A_1 + A_2, \end{aligned}$$

where

$$\tilde{\mathcal{A}}_n(\alpha_0, \beta_0) = \frac{1}{n(n-1)} \sum_{i \neq j} [\pi(\tilde{X}_i) \{1 - \pi(\tilde{X}_j)\} \mathbb{I}(\beta_0^\top \tilde{X}_i > \beta_0^\top \tilde{X}_j)]. \quad (\text{A11})$$

$A_1$  can be expressed as

$$A_1 = \frac{1}{n(n-1)} \sum_{i \neq j} W_{ij} + o_p(1) \quad (\text{A12})$$

with  $W_{ij} = \hat{\pi}(\tilde{X}_i) \{1 - \hat{\pi}(\tilde{X}_j)\} \mathbb{I}(\hat{\beta}^\top \tilde{X}_i > \hat{\beta}^\top \tilde{X}_j) - \pi(\tilde{X}_i) \{1 - \pi(\tilde{X}_j)\} \mathbb{I}(\beta_0^\top \tilde{X}_i > \beta_0^\top \tilde{X}_j)$

By a similar discussion to that for Claim 1, we have  $\hat{\pi}(X_i) \xrightarrow{p} \pi(X_i)$  as  $n \rightarrow \infty$ . In addition, applying Theorem 1 gives  $\hat{\beta}^\top (X_i - X_j) \xrightarrow{p} \beta_0^\top (X_i - X_j)$ . By  $P\{\beta_0^\top (X_i - X_j) = 0\} = 0$  and the continuous mapping theorem (e.g., [33], Theorem 3.2.4), we have that as  $n \rightarrow \infty$ ,

$$E[\mathbb{I}\{\hat{\beta}^\top (X_i - X_j) > 0\}] \xrightarrow{p} E[\mathbb{I}\{\beta_0^\top (X_i - X_j) > 0\}]. \quad (\text{A13})$$

Moreover, similar derivations show that

$$\mathbb{I}\{\widehat{\beta}^\top (X_i - X_j) \leq 0\} \xrightarrow{p} \mathbb{I}\{\beta_0^\top (X_i - X_j) \leq 0\}. \quad (\text{A14})$$

By (A14) and the fact that  $\mathbb{I}(\mathcal{E})\mathbb{I}(\mathcal{E}^c) = 0$  for any event  $\mathcal{E}$ , we have

$$E\left[\mathbb{I}\{\widehat{\beta}^\top (X_i - X_j) > 0\} \mathbb{I}\{\widehat{\beta}^\top (X_i - X_j) \leq 0\}\right] \xrightarrow{p} 0. \quad (\text{A15})$$

Combining (A13) and (A15) yields  $E\left(\left|\frac{1}{n(n-1)} \sum_{i \neq j} W_{ij}\right|\right) = o(1)$ ; thus, by Chebyshev's inequality, we have

$$A_1 \xrightarrow{p} 0 \text{ as } n \rightarrow \infty. \quad (\text{A16})$$

On the other hand, since (A11) is formulated as the U-statistics with bounded kernel, then applying the similar derivation of [34] gives

$$A_2 \xrightarrow{p} 0 \text{ as } n \rightarrow \infty. \quad (\text{A17})$$

Therefore, combining (A16) and (A17) shows that Claim 2 and thus (A10) hold.

**Proof of part (b)**

The proof of part (b) is similar to that of part (a) except for the involvement of the estimates of the function  $H(\cdot)$  and the parameter  $\beta$ . According to Theorem 1, since  $\widehat{\beta}$  and  $\widehat{H}(\cdot)$  are consistent with the true values, then applying a similar argument in part (a) yields the desired result. Thus, the proof is completed.  $\square$

#### Appendix A.5. Proof of Theorem 4

**Proof.** In this appendix, we mainly prove results (a) and (b) separately.

**Proof of part (a)**

Define  $\pi_0(\cdot)$  as (1) with  $\alpha$  being replaced by  $\alpha_0$ . Let  $\widehat{\mathcal{A}}(\alpha_0, \beta_0)$  denote (30) with  $\widehat{\alpha}$  and  $\widehat{\beta}$  being replaced by  $\beta_0$  and  $\beta_0$ , respectively. By adding and subtracting an additional term  $\widehat{\mathcal{A}}(\alpha_0, \beta_0)$ , we have

$$\left\{\widehat{\mathcal{A}}(\widehat{\alpha}, \widehat{\beta}) - \widehat{\mathcal{A}}(\alpha_0, \beta_0)\right\} + \left\{\widehat{\mathcal{A}}(\alpha_0, \beta_0) - \mathcal{A}(\alpha_0, \beta_0)\right\} \triangleq R_1 + R_2. \quad (\text{A18})$$

Note that by applying the Law of Large Numbers and Theorem 1 to the denominator terms of  $R_1$  and  $R_2$ , we have that as  $n \rightarrow \infty$ ,

$$\left\{\frac{1}{n} \sum_{i=1}^n \widehat{\pi}(\widetilde{X}_i)\right\} \left\{\frac{1}{n} \sum_{i=1}^n (1 - \widehat{\pi}(\widetilde{X}_i))\right\} \xrightarrow{p} E\{\pi_0(\widetilde{X}_i)\} E\{1 - \pi_0(\widetilde{X}_i)\}. \quad (\text{A19})$$

To complete this proof, we examine the numerator terms of  $R_1$  and  $R_2$  separately.

##### Step 1: Examine $R_1$

The numerator term of  $R_1$  can be expressed as

$$\begin{aligned} & \frac{1}{n(n-1)} \sum_{i \neq j} \left[ \widehat{\pi}(\widetilde{X}_i) \{1 - \widehat{\pi}(\widetilde{X}_j)\} \widehat{\beta} \mathbb{I}(\widetilde{X}_i > \widetilde{X}_j) - \pi_0(\widetilde{X}_i) \{1 - \pi_0(\widetilde{X}_j)\} \beta_0 \mathbb{I}(\widetilde{X}_i > \widetilde{X}_j) \right] \\ &= \frac{1}{n^2} \sum_{i,j=1}^n \left\{ \widehat{\pi}(\widetilde{X}_i) \widehat{\beta} - \pi_0(\widetilde{X}_i) \beta_0 \right\} \mathbb{I}(\widetilde{X}_i > \widetilde{X}_j) \\ & \quad - \frac{1}{n^2} \sum_{i,j=1}^n \left\{ \widehat{\pi}(\widetilde{X}_i) \widehat{\pi}(\widetilde{X}_j) \widehat{\beta} - \pi_0(\widetilde{X}_i) \pi_0(\widetilde{X}_j) \beta_0 \right\} \mathbb{I}(\widetilde{X}_i > \widetilde{X}_j) \end{aligned} \quad (\text{A20})$$

since  $\mathbb{I}(\tilde{X}_i > \tilde{X}_j) = 0$  if  $i \neq j$ . By the Taylor series expansion, we have

$$\hat{\pi}(\tilde{X}_i)\hat{\beta} - \pi_0(\tilde{X}_i)\beta_0 = \left( \frac{\partial}{\partial \phi} \pi(\tilde{X}_i)\beta \Big|_{\phi=\phi_0} \right) (\hat{\phi} - \phi_0) + o_p(n^{-1/2}),$$

then, the first term in the right-hand side of (A20) gives that

$$\begin{aligned} & \frac{1}{n^2} \sum_{i,j=1}^n \left( \frac{\partial}{\partial \phi} \pi(\tilde{X}_i)\beta \Big|_{\phi=\phi_0} \right) (\hat{\phi} - \phi_0) \mathbb{I}(\tilde{X}_i > \tilde{X}_j) + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n q_i \left( \frac{\partial}{\partial \phi} \pi(\tilde{X}_i)\beta \Big|_{\phi=\phi_0} \right) (\hat{\phi} - \phi_0) + o_p(n^{-1/2}) \\ &= E \left\{ q_i \left( \frac{\partial}{\partial \phi} \pi(\tilde{X}_i)\beta \Big|_{\phi=\phi_0} \right) \right\} (\hat{\phi} - \phi_0) + o_p(n^{-1/2}), \end{aligned} \quad (\text{A21})$$

where  $q_i$  in the second step is defined as  $\frac{1}{n} \sum_{j=1}^n \mathbb{I}(\tilde{X}_i > \tilde{X}_j)$ , and the last step is due to the Law of Large Numbers.

In addition, by the Taylor series expansion, we have

$$\hat{\pi}(\tilde{X}_i)\hat{\pi}(\tilde{X}_j)\hat{\beta} - \pi_0(\tilde{X}_i)\pi_0(\tilde{X}_j)\beta_0 = \left( \frac{\partial}{\partial \phi} \pi(\tilde{X}_i)\pi(\tilde{X}_j)\beta \Big|_{\phi=\phi_0} \right) (\hat{\phi} - \phi_0) + o_p(n^{-1/2}).$$

Then, the second term of (A20) can be expressed as

$$\begin{aligned} & \frac{1}{n^2} \sum_{i,j=1}^n \left\{ \left( \frac{\partial}{\partial \phi} \pi(\tilde{X}_i)\pi(\tilde{X}_j)\beta \Big|_{\phi=\phi_0} \right) \mathbb{I}(\tilde{X}_i > \tilde{X}_j) \right\} (\hat{\phi} - \phi_0) + o_p(n^{-1/2}) \\ &= E \left\{ \left( \frac{\partial}{\partial \phi} \pi(\tilde{X}_i)\pi(\tilde{X}_j)\beta \Big|_{\phi=\phi_0} \right) \mathbb{I}(\tilde{X}_i > \tilde{X}_j) \right\} (\hat{\phi} - \phi_0) + o_p(n^{-1/2}), \end{aligned} \quad (\text{A22})$$

where the equality holds due to the Law of Large Numbers in the U-statistics. Therefore, combining (A21) and (A22) yields

$$\sqrt{n} \left\{ \hat{\mathcal{A}}(\hat{\alpha}, \hat{\beta}) - \hat{\mathcal{A}}(\alpha_0, \beta_0) \right\} = (\mathcal{E}_1 + \mathcal{E}_2) \sqrt{n} (\hat{\phi} - \phi_0) + o_p(1), \quad (\text{A23})$$

where

$$\mathcal{E}_1 = \frac{E \left[ q_i \left\{ \pi_0(\tilde{X}_i) + \frac{\partial}{\partial \alpha} \pi_0(\tilde{X}_i) \beta_0 \right\} \right]}{E \left\{ \pi_0(\tilde{X}_i) \right\} E \left\{ 1 - \pi_0(\tilde{X}_i) \right\}}$$

and

$$\mathcal{E}_2 = \frac{E \left[ \left\{ \pi_0(\tilde{X}_i)\pi_0(\tilde{X}_j) + \frac{\partial}{\partial \alpha} \pi_0(\tilde{X}_i)\pi_0(\tilde{X}_j)\beta_0 \right\} \mathbb{I}(\tilde{X}_i > \tilde{X}_j) \right]}{E \left\{ \pi_0(\tilde{X}_i) \right\} E \left\{ 1 - \pi_0(\tilde{X}_i) \right\}}.$$

## Step 2: Examine $R_2$

Similarly to Step 1, we only examine the numerator term of  $R_2$  because the denominator term comes from (A19).

For the numerator term of  $R_2$ , we have

$$\frac{2}{n(n-1)} \sum_{i \neq j} \left[ \frac{1}{2} \pi_0(\tilde{X}_i) \left\{ 1 - \pi_0(\tilde{X}_j) \right\} \beta_0 \mathbb{I}(\tilde{X}_i > \tilde{X}_j) \right] - E[\pi_0(X_1) \{1 - \pi_0(X_2)\} \beta_0 \mathbb{I}(X_1 > X_2)].$$

$\psi_{ij} \triangleq \frac{1}{2}\pi_0(\tilde{X}_i)\{1 - \pi_0(\tilde{X}_j)\}\beta_0\mathbb{I}(\tilde{X}_i > \tilde{X}_j)$  is the kernel and  $\frac{2}{n(n-1)} \sum_{i \neq j} \psi_{ij}$  forms the U-statistics. By Conditions (C3) and (C4),  $E(\psi_{ij}^2) < \infty$ , and by the construction in Section 3.3.1, we have  $E(\psi_{ij}) = E[\pi_0(X_1)\{1 - \pi_0(X_2)\}\beta_0\mathbb{I}(X_1 > X_2)]$ . Then, applying Theorem 12.3 in [29] and Slutsky's theorem yields that, as  $n \rightarrow \infty$ ,

$$\begin{aligned} & \sqrt{n}\{\hat{\mathcal{A}}(\alpha_0, \beta_0) - \mathcal{A}(\alpha_0, \beta_0)\} \\ &= \frac{1}{E\{\pi_0(\tilde{X}_i)\}E\{1 - \pi_0(\tilde{X}_i)\}} \sqrt{n}\left(\frac{2}{n(n-1)} \sum_{i \neq j} \psi_{ij} - E(\psi_{ij})\right) \\ &\xrightarrow{d} N(0, 2^2\mathcal{E}_3) \end{aligned} \quad (\text{A24})$$

with  $\mathcal{E}_3 = \frac{\text{var}(\psi_{ij})}{[E\{\pi_0(\tilde{X}_i)\}E\{1 - \pi_0(\tilde{X}_i)\}]^2}$ .

Finally, combining (A23) and (A24) gives that as  $n \rightarrow \infty$ ,

$$\sqrt{n}\{\hat{\mathcal{A}}(\hat{\alpha}, \hat{\beta}) - \mathcal{A}(\alpha_0, \beta_0)\} \xrightarrow{d} N(0, \sigma_{\pi})$$

with  $\sigma_{\pi} = (\mathcal{E}_1 + \mathcal{E}_2)^{\top} \Sigma_{\phi} (\mathcal{E}_1 + \mathcal{E}_2) + 4\mathcal{E}_3$  and  $\Sigma_{\phi}$  being the asymptotic variance of  $\sqrt{n}(\hat{\phi} - \phi_0)$  defined after Theorem 2.

**Proof of part (b)**

Following a similar discussion to that in part (a), we consider the decomposition

$$\begin{aligned} \hat{\mathcal{A}}_t(\hat{\alpha}, \hat{\beta}) - \mathcal{A}_t(\alpha_0, \beta_0) &= \{\hat{\mathcal{A}}_t(\hat{\alpha}, \hat{\beta}) - \hat{\mathcal{A}}_t(\alpha_0, \beta_0)\} + \{\hat{\mathcal{A}}_t(\alpha_0, \beta_0) - \mathcal{A}_t(\alpha_0, \beta_0)\} \\ &= K_1 + K_2. \end{aligned} \quad (\text{A25})$$

By applying the Law of Large Numbers and Theorem 1 to the denominator terms of  $K_1$  and  $K_2$ , we have that as  $n \rightarrow \infty$ ,

$$\begin{aligned} & \left[ \frac{1}{n} \sum_{i=1}^n \hat{\pi}(\tilde{X}_i) \{1 - \hat{S}_u(t|\tilde{X}_i)\} \right] \left[ \frac{1}{n} \sum_{i=1}^n \{\hat{\pi}(\tilde{X}_i) \hat{S}_u(t|\tilde{X}_i)\} \right] \\ &\xrightarrow{p} E[\pi_0(\tilde{X}_i) \{1 - S_u(t|\tilde{X}_i)\}] E\{\pi_0(\tilde{X}_i) S_u(t|\tilde{X}_i)\}. \end{aligned} \quad (\text{A26})$$

On the other hand, the numerator term of  $K_1$  can be expressed as

$$\begin{aligned} & \frac{1}{n(n-1)} \sum_{i \neq j} [\hat{\pi}(\tilde{X}_i) \{1 - \hat{S}_u(t|\tilde{X}_i)\} \hat{\pi}(\tilde{X}_j) \hat{S}_u(t|\tilde{X}_j) \mathbb{I}(\hat{\beta}^{\top} \tilde{X}_i > \hat{\beta}^{\top} \tilde{X}_j)] \\ & - \frac{1}{n(n-1)} \sum_{i \neq j} [\pi_0(\tilde{X}_i) \{1 - S_u(t|\tilde{X}_i)\} \pi_0(\tilde{X}_j) S_u(t|\tilde{X}_j) \mathbb{I}(\beta^{\top} \tilde{X}_i > \beta^{\top} \tilde{X}_j)] \\ &= \frac{1}{n^2} \sum_{i,j=1}^n [\hat{\pi}(\tilde{X}_i) \{1 - \hat{S}_u(t|\tilde{X}_i)\} \hat{\pi}(\tilde{X}_j) \hat{S}_u(t|\tilde{X}_j) \mathbb{I}(\hat{\beta}^{\top} \tilde{X}_i > \hat{\beta}^{\top} \tilde{X}_j)] \\ & - \frac{1}{n^2} \sum_{i,j=1}^n [\pi_0(\tilde{X}_i) \{1 - S_u(t|\tilde{X}_i)\} \pi_0(\tilde{X}_j) S_u(t|\tilde{X}_j) \mathbb{I}(\beta^{\top} \tilde{X}_i > \beta^{\top} \tilde{X}_j)] \\ &= \frac{1}{n} \sum_{i=1}^n q_i \left[ \frac{\partial}{\partial \phi} \pi(\tilde{X}_i) \{1 - S_u(t|\tilde{X}_i)\} \pi(\tilde{X}_j) S_u(t|\tilde{X}_j) \right]_{\phi=\phi_0} (\hat{\phi} - \phi_0) + o_p(1). \end{aligned} \quad (\text{A27})$$

In addition, the numerator term of  $K_2$  can be written as

$$\frac{2}{n(n-1)} \sum_{i \neq j} \Gamma_{ij} - E(\Gamma_{ij}), \quad (\text{A28})$$

where  $\Gamma_{ij} = \frac{1}{2}\pi_0(\tilde{X}_i)\{1 - S_u(t|\tilde{X}_i)\}\pi_0(\tilde{X}_j)S_u(t|\tilde{X}_j)\mathbb{I}(\beta^\top \tilde{X}_i > \beta^\top \tilde{X}_j)$  is the kernel of the U-statistic in the numerator term of  $K_2$ .

Finally, combining (A25)–(A28) with the central limit theorem in  $\hat{\phi} - \phi_0$  and  $\Gamma_{ij}$  and using Slutsky's theorem give that as  $n \rightarrow \infty$ ,

$$\sqrt{n}\{\hat{\mathcal{A}}_t(\hat{\alpha}, \hat{\beta}) - \mathcal{A}_t(\alpha_0, \beta_0)\} \xrightarrow{d} N(0, \sigma_{\text{TD}}),$$

where  $\sigma_{\text{TD}} = \mathcal{M}_1^\top \Sigma_\phi \mathcal{M}_1 + 4\mathcal{M}_2$ , with  $\mathcal{M}_2 = \frac{\text{var}(\Gamma_{ij})}{E[\pi_0(\tilde{X}_i)\{1 - S_u(t|\tilde{X}_i)\}]E\{\pi_0(\tilde{X}_i)S_u(t|\tilde{X}_i)\}}$  and  $\mathcal{M}_1 = E\left(q_i \left[ \frac{\partial}{\partial \phi} \pi(\tilde{X}_i) \{1 - S_u(t|\tilde{X}_i)\} \pi(\tilde{X}_j) S_u(t|\tilde{X}_j) \right]_{\phi=\phi_0} \right)$ .  $\square$

## References

1. Lawless, J.F. *Statistical Models and Methods for Lifetime Data*; Wiley: New York, NY, USA, 2003.
2. Lu, W.; Ying, Z. On semiparametric transformation cure models. *Biometrika* **2004**, *91*, 331–343. [\[CrossRef\]](#)
3. Chen, C.-M.; Shen, P.-S.; Wei, J.C.-C.; Lin, L. A semiparametric mixture cure survival model for left-truncated and right-censored data. *Biom. J.* **2017**, *59*, 270–290. [\[CrossRef\]](#)
4. Amico, M.; Keilegom, I.V.; Legrand, C. The single-index/Cox mixture cure model. *Biometrics* **2019**, *75*, 452–462. [\[CrossRef\]](#)
5. [\[CrossRef\]](#) Amico, M.; Keilegom, I.V. Cure models in survival analysis. *Annu. Rev. Stat. Its Appl.* **2018**, *5*, 311–342. [\[CrossRef\]](#)
6. Bertrand, A.; Legrand, C.; Carroll, R.J.; Meester, C.D.; Keilegom, I.V. Inference in a survival cure model with mismeasured covariates using a simulation-extrapolation approach. *Biometrika* **2017**, *104*, 31–50. [\[CrossRef\]](#)
7. Ma, Y.; Yin, G. Cure rate model with mismeasured covariates under transformation. *J. Am. Stat. Assoc.* **2008**, *103*, 743–756. [\[CrossRef\]](#)
8. Chen, L.-P. Semiparametric estimation for cure survival model with left-truncated and right-censored data and covariate measurement error. *Stat. Probab. Lett.* **2019**, *154*, 108547. [\[CrossRef\]](#)
9. Beyene, K.M.; El Gouch, A. Smoothed time-dependent receiver operating characteristic curve for right censored survival data. *Stat. Med.* **2020**, *39*, 3373–3396. [\[CrossRef\]](#)
10. Heagerty, P.J.; Lumley, T.; Pepe, M.S. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **2000**, *56*, 337–344. [\[CrossRef\]](#)
11. Kamarudin, A.N.; Cox, T.; Kolamunnage-Dona, R. Time-dependent ROC curve analysis in medical research: Current methods and applications. *BMC Med. Res. Methodol.* **2017**, *17*, 53. [\[CrossRef\]](#)
12. Li, L.; Greene, T.; Hu, B. A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data. *Stat. Methods Med. Res.* **2018**, *27*, 2264–2278. [\[CrossRef\]](#)
13. Song, X.; Zhou, X.-H. A semiparametric approach for the covariate specific roc curve with survival outcome. *Stat. Sin.* **2008**, *18*, 947–965.
14. Zhang, Y.; Han, X.; Shao, Y. The ROC of Cox proportional hazards cure models with application in cancer studies. *Lifetime Data Anal.* **2021**, *27*, 195–215. [\[CrossRef\]](#)
15. Amico, M.; Keilegom, I.V.; Han, B. Assessing cure status prediction from survival data using receiver operating characteristic curves. *Biometrika* **2021**, *108*, 727–740. [\[CrossRef\]](#)
16. Kolamunnage-Dona, R.; Kamarudin, A.N. Adjustment for the measurement error in evaluating biomarker performances at baseline for future survival outcomes: Time-dependent receiver operating characteristic curve within a joint modelling framework. *Res. Methods Med. Health Sci.* **2021**, *2*, 51–60. [\[CrossRef\]](#)
17. Crowther, M.J.; Lambert, P.C.; Abrams, K.R. Adjusting for measurement error in baseline prognostic biomarkers included in a time-to-event analysis: A joint modelling approach. *BMC Med. Res. Methodol.* **2013**, *13*, 146. [\[CrossRef\]](#)
18. Nevo, D.; Zucker, D.M.; Tamimic, R.M.; Wang, M. Accounting for measurement error in biomarker data and misclassification of subtypes. *Stat. Med.* **2016**, *35*, 5686–5700. [\[CrossRef\]](#)
19. Mao, M.; Wang, J.-L. Semiparametric efficient estimation for a class of generalized proportional odds cure models. *J. Am. Stat.* **2010**, *105*, 302–311. [\[CrossRef\]](#)
20. Chen, L.-P. Semiparametric estimation for the transformation model with length-biased data and covariate measurement error. *J. Stat. Comput. Simul.* **2020**, *90*, 420–442. [\[CrossRef\]](#)
21. Carroll, R.J.; Ruppert, D.; Stefanski, L.A.; Crainiceanu, C.M. *Measurement Error in Nonlinear Model*; CRC Press: New York, NY, USA, 2006.
22. Chen, L.-P.; Yi, G.Y. Semiparametric estimation methods for left-truncated and right-censored survival data with covariate measurement error. *Ann. Inst. Stat. Math.* **2021**, *73*, 481–517. [\[CrossRef\]](#)

23. Elandt-Johnson, R.C.; Johnson, N.L. *Survival Models and Data Analysis*; John Wiley & Sons: New York, NY, USA, 1980.
24. Chen, L.-P.; Yi, G.Y. Analysis of noisy survival data with graphical proportional hazards measurement error models. *Biometrics* **2021b**, *77*, 956–969. [[CrossRef](#)]
25. Hunter, D.R.; Lange, K. Computing estimates in the proportional odds model. *Ann. Inst. Stat. Math.* **2002**, *54*, 155–168. [[CrossRef](#)]
26. Zheng, Y.; Cai, T.; Feng, Z. Application of the time-dependent ROC curves for prognostic accuracy with multiple biomarkers. *Biometrics* **2006**, *62*, 279–287. [[CrossRef](#)]
27. Carroll, R.J.; Spiegelman, C.H.; Lan, K.K.G.; Bailey, K.T.; Abbot, R.D. On errors-in-variables for binary regression model. *Biometrika* **1984**, *71*, 19–25. [[CrossRef](#)]
28. Webb, A.; Ma, J. Cox models with time-varying covariates and partly-interval censoring—A maximum penalised likelihood approach. *Stat. Med.* **2023**, *42*, 815–833. [[CrossRef](#)]
29. van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press: New York, NY, USA, 1998.
30. Kim, J.P.; Lu, W.; Sit, T.; Ying, Z. A unified approach to semiparametric transformation models under general biased sampling schemes. *J. Am. Stat. Assoc.* **2013**, *108*, 217–227. [[CrossRef](#)]
31. van der Vaart, A.W.; Wellner, J.A. *Weak Convergence and Empirical Processes*; Springer: New York, NY, USA, 1996.
32. Su, Y.-R.; Wang, J.-L. Modeling left-truncated and right-censored survival data with longitudinal covariates. *Ann. Stat.* **2012**, *40*, 1465–1488. [[CrossRef](#)]
33. Durrett, R. *Probability: Theory and Examples*; Cambridge University Press: New York, NY, USA, 2010.
34. Hoeffding, W. A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **1948**, *19*, 293–325. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.