Université catholique de Louvain
Institute of Statistics, Biostatistics and Actuarial Sciences

Time-Dependent ROC Curve Estimation and Inference for Censored Data:

Some Novel Contributions

Kassu Mehari BEYENE

A thesis submitted to the
Université catholique de Louvain
in partial fulfillment of the
requirements for the degree of
Doctor of Sciences

Thesis committee:

Prof. Anouar El Ghouch	UCLouvain	Supervisor
Prof. Catherine Legrand	UCLouvain	Chairperson
Prof. Ingrid Van Keilegom	UCLouvain & KULeuven	
Prof. Abderrahim Oulhaj	UAE University	
Prof. Juan Carlos Pardo-Fernández	University of Vigo	

September 2020

Acknowledgments

It gives me a great pleasure to come to this final point to express my deepest gratitude to all individuals who contributed to this PhD thesis and supported me in one way or another during this journey, which would not have been possible without any of them.

I would like to express my heartfelt gratitude to my supervisor, Professor Anouar El Ghouch, for his sincere, faithful and immense devotion to help me for the accomplishment of this thesis work and to bring me here from the start. It has been an honour and a privilege working with you. Anouar, thank you so much for the freedom you gave me to come to your office, to ask you questions, and to discuss difficulties at any time. This work certainly would not have been possible without your support and guidance. Anouar, you are not only a great supervisor, but you are also a great human being who always dedicated to supporting me in every possible way. I will forever be grateful to all your incredible supports. I would also like to extend my gratitude and appreciation to the jury members, Professor Catherine Legrand, Professor Ingrid Van Keilegom, Professor Abderrahim Oulhaj, and Professor Juan Carlos Pardo-Fernández for the time that they have taken to read the thesis and for making constructive comments and remarks which have improved the thesis. Catherine, Ingrid and Abderrahim thank you so much for following my progress during the last four years and for your interest in my work. Abderrahim, thank you so much for the discussions we had about the receiver operating characteristic (ROC) curve and for the collaboration on the ROC curve with cure fraction and for your help on writing the first paper. Juan, thank you so much for your enthusiasm during the private defense.

Many thanks to my colleagues at ISBA and SMCS: Alexandre, Joris, Ensiyeh,

Xavier, Hassana, Stefka, Mickaël, Nathalie, Oswaldo, Rebecca, Sophie, Fadoua, Pauline, Charles-Guy, Emmanuel, Antoine, Vanessa, Keivan, John-John, Gilles and Peter. Special thanks to my office mates: Joris, Ensiyeh, Xavier and Maïlis for the discussions and sharing information. I also thank the whole administrative staff at ISBA: Nancy Guillaume, Sophie Malali, Nadja Peiffer and Tatiana Regout. Nancy, I am very much thankful for all your support.

I am also indebted to my friends: Alemu, Worku, Yimer, Negera, Jamal, Mohammed, Yemisrach, Solomon, Haymanot, Abraham and many others for asking me about my study progress and for the time we had together. Alexo, thank you so much not only for all your help and useful suggestions but also for being there to listen when I needed an ear. To *Tswa* program members, thank you so much for praying together. May GOD bless you all.

I owe a special thanks to an extraordinary person, my wife Sara Abera, for her continued and unfailing support and understanding of my goals and aspirations. Without her supports, I would not have been able to complete much of what I have done. I greatly value her contribution and highly appreciate her belief in me. Sara, your patience and sacrifice will remain my inspiration throughout my life. My son Kalab also has been my happiness and encouragements during the last six years of my studies. Words would never say how grateful I am to both of you. I am the luckiest in the world to have such a lovely and caring family, standing beside me with their love and unconditional support.

Above all, I praise and thank the ALMIGHTY GOD, the merciful and the lord of all the worlds, for blessing me with countless opportunities and for surrounding me with many great people. I will always remain grateful to you.

Kassu Mehari Beyene

Louvain la Neuve, September 2020

Summary

In recent years, prediction models have become increasingly popular tools to estimate the risk of a person developing a specific event of interest at a particular time, given his/her characteristics. Accuracy of these predictive models is critical as it determines the quality of their predictions that form scientific evidence for informing treatment or other clinical decisions for individual patients. Therefore, nowadays, it is widely accepted that assessing the predictive accuracy of predictive models is a critical step before make use of them for a clinical practice. The ability of the models to distinguish between two classes (e.g., case and control subjects) is one of the most important criteria to evaluate predictive accuracy. To this end, the receiver operating characteristic (ROC) curve and the associated summary indices are the most widely used tools to evaluate the classification accuracy of the model. There exist various methods to estimate the time-dependent ROC curve and the associated summary indices. However, there is a lack of literature that addresses issues such as the presence of cure fraction, interval censoring, and smoothing the ROC curve. This thesis fills these gaps by introducing some new estimation and inference methods for time-dependent ROC curves and associated summary measures.

Chapter 1 is an introductory chapter that presents an overview of the basic concepts of survival data analysis. In particular, this chapter starts by introducing functions that are important to characterize survival data and discus about the peculiar features of survival data. Also, it highlights some of the estimation methods in the context of censored survival data. The chapter then proceeds to introduce concepts and techniques of modeling survival data with cure fraction. Finally, before closing it with the organization of the thesis, this chapter provides a brief introduction about predictive accuracy evaluation and the bootstrap approach.

Chapter 2 studies the effects of violating the assumption behind the classical survival models on the validity of prediction accuracy measures. More specifically, this chapter first investigates the validity of a non-parametric time-dependent area under the ROC curve (AUC) estimator of the classical right-censored data, proposed by Li et al. (2018), in the presence of cure fraction. To this end, it introduces definitions for the time-dependent true-positive and false-positive rates in the presence of cure fraction and show their equivalency with that of Heagerty et al. (2000). Second, motivated by the current practice of deriving risk tools from survival models that do not acknowledge the presence of cure, the chapter also examines the loss, in terms of discriminatory performance, of using a classical survival model to derive risk tools. The finding of a comprehensive simulation study attests to the validity of the studied estimator. Moreover, the study also exhibited that the risk score derived from the classical survival model has less discriminatory ability than the risk score derived from the cure model. An illustrative application to a real survival data with cure fraction also supports the findings of the simulation study.

The third chapter introduces some non-parametric methods to estimate time-dependent ROC curves for the right censored data. This chapter first provides an empirical (non-smoothed) estimator of the time-dependent ROC curves and, afterward, a smooth estimator using the well-known kernel techniques. This latter estimator, however, suffers from a boundary problem because it has bounded support in the unit interval $[0, 1]$. To overcome the boundary problem, the proposed smooth ROC curve estimator is then polished using a quantile transformation technique. This chapter also provides some data-driven bandwidth selection methods for the bandwidth parameter needed to smooth the ROC curve, and it introduces a data-based bootstrap resampling method to estimate variability and make inference. The finite sample performance of the proposed time-dependent ROC curve estimators and bandwidth selection methods are illustrated through an extensive simulation study. Also, the robustness of the proposed ROC estimator to the choice of smoothing parameters and the validity of the bootstrap approach are shown via a simulation study. Finally, the application of the proposed estimators is illustrated using a real data example.

Chapter 4 presents the extensions of methods discussed in Chapter 3 for the case of general censored data, specifically interval censored data. The time-dependent ROC curve estimation methods proposed in Chapter 3, however, cannot accommodate interval censored survival data in which case the survival function estimators are not valid. Therefore, using a non-parametric, semi-parametric, and parametric methods for the survival function estimation, this chapter offers an empirical (non-smoothed) and a smoothed time-dependent ROC curve estimation methods. To summarize the ROC curve, the two commonly used summary measures: the time-dependent AUC and the time-dependent Youden index estimators for interval

censored data are proposed. A non-parametric data-based bootstrap approach is suggested to estimate the variability and construct confidence intervals for the ROC curve and the associated summary measures. The results of a thorough simulation study are presented to investigate the finite sample performance of the proposed methods. An application to two real data sets from the literature also illustrates the use of the proposed methods.

Chapter 5 discusses the results and highlight the main contributions of the thesis. The remainder of this chapter, based on the findings of this thesis, presents possible directions of potentially rewarding future research(es) to be carried out.

Contents

Summary **v**

1 Introduction **1**
1.1 Introduction to survival data analysis 2
1.1.1 Basic functions in survival analysis 3
1.1.2 Features of survival data 4
1.2 Survival estimation 5
1.2.1 Nonparametric estimation 6
1.2.2 Regression modeling in survival analysis 10
1.3 Survival estimation in the presence of cure fraction 13
1.3.1 Mixture cure model 13
1.3.2 Promotion time cure model 14
1.4 Evaluating prediction accuracy 15
1.4.1 The receiver operating characteristic (ROC) curve 16
1.4.2 Some ROC curve summary measures 19
1.5 The bootstrap approach 23
1.5.1 Bootstrapping 24
1.5.2 Bootstrap confidence intervals 25
1.6 Outline of the thesis 25

2 On the Validity of Time-Dependent AUC Estimation in the Presence of Cure Fraction **27**
2.1 Introduction 28
2.2 Method 30
2.2.1 The estimation of ROC curve and AUC in survival models: The nonparametric simple method 31

2.2.2 Derivation of prognostic tools in survival analysis 33
2.2.3 Survival models with cure 34
2.3 Simulations 35
2.3.1 The validity of the NPS estimator under cure: The case of a known prognostic tool 35
2.3.2 The validity of the NPS estimator under cure: The case of an estimated prognostic tool 39
2.3.3 Deriving risk tools from survival models that ignore the presence of cure 47
2.4 Predicting Node-Negative Breast Cancer 49
2.5 Discussion and conclusions 52

3 Smoothed Time-Dependent ROC Curve for Right Censored Survival Data 55
3.1 Introduction 56
3.2 Estimation method 58
3.2.1 Imputation 59
3.2.2 Estimation 59
3.2.3 Some asymptotic results 61
3.2.4 Boundary correction 62
3.2.5 Variance estimation and inference 64
3.3 Bandwidth selection 64
3.4 Simulation Study 68
3.4.1 Data generation 68
3.4.2 Simulation results 71
3.4.3 Comparing the proposed method with the competitors 77
3.4.4 Validity of the bootstrap 80
3.5 Real data analysis 82
3.6 Discussion 85

4 Time-Dependent ROC Curve Estimation for Interval Censored Data 87
4.1 Introduction 88
4.2 Method 89
4.2.1 Definitions 90
4.2.2 ROC estimation 90
4.2.3 Conditional survival function estimation 93
4.2.4 Bootstrap and inference 94
4.3 Simulation study 95
4.3.1 Data generation 95
4.3.2 Simulation results for the ROC 97
4.3.3 Summary measures estimators and the bootstrap 102
4.4 Real data application 105

4.4.1 The Hypobaric Decompression Sickness (HDS) data 105
4.4.2 The Signal Tandmobiel® (ST) data 107
4.5 Discussion 109

5 Discussion and Future Research Directions 113
5.1 Discussion and general conclusion 114
5.2 Future research directions 117

References 121

A The cenROC Package 133

List of Tables

1.1 Some commonly used parametric distributions 4

2.1 Parameters used to generate the data for Scenario 1 and 2 and the resulting percentages of cure (%cure) and true values of $AUC_M(t)$. . 37
2.2 % bias and MSE of the NPS AUC estimate for data generated according to Scenario 1 and 2. 38
2.3 Parameters used to generate the data for case 1 and case 2 and the resulting percentages of cure (%cure) and true values of $AUC_{M^{\mathcal{MCM}}}(t)$. 41
2.4 The percent bias and MSE for Case 1 and Case 2 (Approach 1) MCM is mixture cure model and COX is classical Cox model. 44
2.5 The percent bias and MSE for MCM-case 2 (Approach 2). 47

3.1 MIB($\times 10^{-2}$) and MISE($\times 10^{-3}$) of the ROC estimators with and without boundary correction for different sample sizes (n), censoring rates (% cen), p and t values obtained with the **first scenario** using the three bandwidth selection methods: NR, PI and CV. 72
3.2 MIB($\times 10^{-2}$) and MISE($\times 10^{-3}$) of the ROC estimators with and without boundary correction for different sample sizes (n), censoring rates (% cen), p and t values obtained with the **second scenario** using the three bandwidth selection methods: NR, PI and CV. . . . 73
3.3 The MISE($\times 10^{-3}$) of ROC estimators with and without boundary correction computed with different censoring rates (% cen), $\boldsymbol{n = 100}$, $p = 0.25$ and $t = Q_2 = 1$ for the first scenario and with $b = c_1 \times n^{-1/5}$ and $h = c_2 \times n^{-1/3}$. The last two columns correspond to the NR and the PI selection methods for h and the last row ($\hat{b}$) corresponds to Sheather and Jones's selection methods for b. 75

3.4 The MISE($\times 10^{-3}$) of ROC estimators with and without boundary correction computed with different censoring rates (% cen), $\boldsymbol{n = 400}$, $p = 0.25$ and $t = Q_2 = 1$ for the first scenario and with $b = c_1 \times n^{-1/5}$ and $h = c_2 \times n^{-1/3}$. The last two columns correspond to the NR and the PI selection methods for h and the last row ($\hat{b}$) corresponds to Sheather and Jones's selection methods for b. 76

3.5 The MIB($\times 10^{-2}$) and MISE($\times 10^{-3}$) of the empirical, the smooth (boundary corrected), and the MP estimators computed for different sample sizes (n), censoring rates (% cen), p and t values obtained with the **first scenario**. The bandwidth(s) are selected using the NR method for our estimator and the LSCV and the PI methods for the MP estimator. 78

3.6 The MIB($\times 10^{-2}$) and MISE($\times 10^{-3}$) of the empirical, the smooth (boundary corrected), and the MP estimators computed for different sample sizes (n), censoring rates (% cen), p and t values obtained with the **second scenario**. The bandwidth(s) are selected using the NR method for our estimator and the LSCV and the PI methods for the MP estimator. 79

3.7 The empirical standard deviation (ESD), the average bootstrap standard deviation (ASD_B), the coverage probability (CP) and the average width (AW) of 95%−bootstrap confidence intervals obtained with the smooth estimator using the standard normal method (Normal) and the percentile method computed for different sample sizes (n), censoring rates (% cen), p and t values. Data simulated according to the **first scenario.** 81

3.8 The estimated AUC value with 95% bootstrap confidence intervals using the percentile method and the standard normal method (Normal) from both the proposed smoothed and empirical (Non-smoothed) methods for $t = 3$, 6 years. 83

3.9 The estimated Youden index and optimal cut-off point obtained from the smoothed and non-smoothed estimators for $t = 3$, 6 years. 83

3.10 The true and estimated Youden index, optimal cut-off point, sensitivity and specificity obtained from a data simulated according to the **first scenario** ($t = 0.3$ and $p = 0.25$). 85

4.1 The MIB($\times 10^{-2}$) and MISE($\times 10^{-2}$) of the proposed empirical ROC estimator and the WU method computed for different sample sizes (n), average intervals lengths (AL), right censoring rates (%RC), and t values. **Case I,** $\rho = 0.5$. 99

4.2 The MIB($\times 10^{-2}$) and MISE($\times 10^{-2}$) of the proposed empirical ROC estimators and the WU's method computed for different left censoring rates (%LC), sample sizes (n), ρ and t values. **Case II,** $AL = 0.75$.100

4.3 The MIB($\times 10^{-2}$) and MISE($\times 10^{-2}$) of the proposed empirical and smooth ROC estimators for different sample sizes (n), average intervals lengths (AL), right censoring rates (%RC), and t values. **Case I,** $\rho = 0.5$. 101
4.4 Bias ($\times 10^{-2}$) and MSE ($\times 10^{-2}$) of the proposed time-dependent AUC estimators and the WU method computed for different right censoring rates (%RC), average intervals lengths (AL), sample sizes (n), ρ and t values. **Case I**. 103
4.5 Bias ($\times 10^{-2}$) and MSE ($\times 10^{-2}$) for the empirical (non-smoothed) and smoothed Youden index estimators computed for different right censoring rates (%RC), average intervals lengths (AL), sample sizes (n), ρ and t values and for the corresponding optimal cut-point bias ($\times 10^{-1}$) and MSE ($\times 10^{-1}$). **Case I**. 104
4.6 Coverage probabilities (CP) and the average widths (AW) of 95% confidence intervals obtained with the bootstrap using the standard or the percentile method. These results correspond to different sample sizes (n), right censoring rates (%RC), interval lengths (AL) and p values. **Case I,** $t = Q_2 = 1.8$. 105
4.7 Cox proportional hazards regression estimates for the HDS data . . 106
4.8 Estimated AUC with the associated 95% percentile bootstrap confidence intervals (CI) from the HDS data with $t = 1$, 2, and 3 years 107
4.9 Estimated Youden index with the associated 95% percentile bootstrap confidence intervals (CI) from the HDS data with $t = 1$, 2, and 3 years. 107
4.10 Estimated Youden index based optimal cut-off values from the HDS data with $t = 1$, 2, and 3 years. 107

List of Figures

1.1 Plot to illustrate the three censoring mechanism; the solid diamonds denote the unobserved times when the event of interest occurred. . . 5
1.2 Hypothetical marker distributions in healthy and diseased populations and the corresponding ROC curves. 22
1.3 A ROC curve with the Youden index and the corresponding cutoff value. 23

2.1 Boxplots of $\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)$ in Case 1 for the time horizon $t = 2$. The solid circles and triangles indicates the "true" AUC values $AUC_{M^{\mathcal{W}}}(t)$ of the Cox and mixture cure working models, respectively. 42
2.2 Boxplots of $\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)$ in Case 2 for the time horizon $t = 2$. The solid circles and triangles indicates the "true" AUC values $AUC_{M^{\mathcal{W}}}(t)$ of the Cox and mixture cure working models, respectively. 43
2.3 The boxplots of the averaged NPS AUC estimates $\hat{AAUC}_{\hat{M}_k^{\mathcal{W}}}$ obtained with MCM-case 2 for a time horizon $t = 2$. The solid circles and triangles indicate the true or the population version of the AUC ($AUC_{\hat{M}_k^{\mathcal{W}}}$) of the Cox and mixture cure working models, respectively. 46
2.4 The boxplots of $AUC_{\hat{M}_k^{\mathcal{W}}}$, the population version of AUC of the prognostic tool $\hat{M}_k^{\mathcal{W}}$ obtained with MCM-case 2 for a time horizon $t = 2$ with $\mathcal{W} = MCM, COX$. 49
2.5 Kaplan-Meier survival estimate of the lymph-node-negative breast cancer patients (+ : censored observations). 51

2.6 ROC curve and the corresponding AUC estimates for the MCM (solid line) and Cox model (dashed line) obtained for $t = 1$ year and using all data as training and validation set. 51
2.7 ROC curve and the corresponding AUC estimates for the MCM (solid line) and Cox model (dashed line) obtained for $t = 1$ year and by splitting the data to training and validation set. 51
2.8 ROC curve and the corresponding AUC estimates for the MCM (solid line) and Cox model (dashed line) obtained for $t = 3$ years and using all data as training and validation set. 52
2.9 ROC curve and the corresponding AUC estimates for the MCM (solid line) and Cox model (dashed line) obtained for $t = 3$ years and by splitting the data to training and validation set. 52

3.1 The true ROC curves and the corresponding AUC values of the first scenario (left column) and the second scenario (right column) for the prediction time Q_1 (solid line), Q_2 (dashed line) and Q_3 (doted line). 70
3.2 Estimated time-dependent ROC curve from the empirical (solid line), boundary corrected smooth (dashed line) and the MP method (doted line) for 3 years (left) and 6 years (right) prediction time. 83
3.3 The estimated time-dependent ROC curve together with the estimated Youden index and the associated optimal cut-off point at $t = 370$ with (left) and without (right) smoothing for the ovarian cancer survival data. 85

4.1 The true ROC curve and the corresponding AUC and Youden function for $\rho = 0.25$ (top row) and for $\rho = 0.5$ (bottom row) for the prediction time Q_1 (solid line), Q_2 (dashed line) and Q_3 (dotted line). 97
4.2 The estimated ROC curves and the associated AUC values for boys (top row) and for girls (bottom row) with six years (left panel) and nine years (right panel) prediction time. Spline based method: solid line; AFT-Weibull based method: dashed line; WU method: dotted line. 109

A.1 Estimated time-dependent ROC curve with 95% confidence band obtained using percentile bootstrap approach. 134
A.2 Estimated time-dependent ROC curve for HDS data. 135
A.3 Estimated time-dependent ROC curve with the Youden index and the corresponding optimal cutoff point. 136

CHAPTER 1

Introduction

In clinical medicine, identifying individuals with high risk of developing a certain event of interest is crucial as it helps to plan early prevention and possibly treatment. The event can be death due to certain disease, diagnosis of a disease or the recurrences of a disease. In particular, the use of prediction models to predict the risk of developing a disease (e.g., diabetes, breast cancer, asthma, dementia, ...), given the individuals characteristics and diagnostic test(s) measures, is very popular. Nowadays, prognostic risk scores (or markers) derived from statistical models are widely used to identify high risk individuals. To this end, survival analysis plays an important role when the outcome of interest is time to an event. In public health, for example, the Framingham risk score is a well-known tool used to predict the probability of developing cardiovascular disease within 10-year period. This probability is estimated from a risk score obtained by combining risk factors such as age, blood pressure, smoking status, diabetes, high density lipoproteins and total cholesterol levels. However, before approving the use of a prognostic risk score in routine clinical practice, its predictive quality must be rigorously evaluated through statistical analysis. By taking a model score as the diagnostic marker, the receiver operating characteristic (ROC) curves based classification accuracy measure is the most popular method to evaluate the predictive ability of a binary outcome model. Over the past few decades, several extensions of the standard binary classification accuracy measures to the context of time-to-event outcomes have been proposed; see, for example, Etzioni et al. (1999); Heagerty et al. (2000);

Slate and Turnbull (2000); Heagerty and Zheng (2005). One of the challenges of the time-dependent ROC curve estimation is that the event time is often not observable due to the presence of censoring. To deal with this issue, in the literature, various time-dependent ROC curves estimation methods have been proposed (see, for example, Heagerty et al. (2000); Heagerty and Zheng (2005); Martínez-Camblor et al. (2016); Martínez-Camblor and Pardo-Fernández (2018); Li et al. (2018); Li and Ma (2011); Chambless and Diao (2006) and the references given therein). The main aim of this thesis is to propose some novel time-dependent ROC curve and associated summary measures estimation and inference methods for censored survival data.

The remainder of this introductory chapter is devoted to introduce some basic concepts about survival analysis and prediction accuracy measures. These include notations and methods on which the methodologies proposed in this thesis are based. More precisely, Section 1.1 introduces the concept of survival data, including definitions of basic functions for censored survival data and what distinguishes survival analysis from other statistical methods. Then, Section 1.2 discusses methods for estimating survival quantities. Section 1.3 introduces the concept of cure fraction and describes some important classes of cure models. The concept of measuring the predictive accuracy of the survival model is introduced in Section 1.4. A short introduction about the bootstrap approach is presented in Section 1.5. Finally, Section 1.6 concludes this chapter with the general overview of the outline about the contributions of this thesis.

1.1 Introduction to survival data analysis

Survival analysis, or more generally, time-to-event analysis, refers to a collection of statistical methods for data analysis, for which the outcome variable of interest is the time until an event occurs. The origin of survival analysis dates back to the seventeenth century when the first-ever lifetable was published by Jhone Graunt in 1662 (Kreager; 1988). As the name indicates, the original event of interest was death. Therefore, throughout the centuries, it was linked to the analysis of mortality data. Now the scope of the survival analysis has become broad because of the newly developed statistical methods. Today it becomes the most prominent method for analyzing data from different fields such as medicine, sociology, engineering, marketing, economics, and so on. As a result, the event of interest is no longer limited to time to death, but now it includes time to recurrence, time to progression, time to get a new job, time to machine failure, time to get a second child, and so on. Depending on the area of application, other terms such as "failure time analysis", "event history analysis", "duration analysis" and "reliability analysis" are also used to refer survival analysis.

The modern era in the field of survival analysis starts in the mid-twentieth cen-

tury. This new era was inaugurated by the path-breaking work by Kaplan and Meier (1958), where they published their famous nonparametric method of survival function estimation for right censored time-to-event data. The introduction of the proportional hazards model by David Cox in 1972 (Cox; 1972) was another profound contribution that revolutionized survival analysis. These are two of the most cited papers in the history of statistics with more than 50,000 citations each. Following these publications, the literature has risen rapidly with major advances in both theoretical and applied researches. The comprehensive overview of survival methods with their software applications are available in different books including (but not limited to) Kalbfleisch and Prentice (2011), Collett (2015), Tableman and Kim (2003) and Bogaerts et al. (2017).

1.1.1 Basic functions in survival analysis

Survival time is a random variable that measures the time to occurrence of an event of interest. The distribution of survival times is commonly characterized by three functions: the survival function, the hazard function, and the cumulative hazard function. Let T denote a continuous, nonnegative random variable representing the event time of interest with probability density function $f(t)$ and the corresponding cumulative distribution function $F(t) = P(T \leq t) = \int_0^t f(u)du$. The survival function of T, denoted by $S(t)$, is defined as the probability that the event of interest has not yet occurred by time t. It is mathematically defined as

$$S(t) = P(T > t) = 1 - F(t).$$

Another important function for describing the survival time is the hazard function $h(t)$, defined as

$$h(t) = \lim_{\Delta \to 0} \frac{P(t < T < t + \Delta | T \geq t)}{\Delta}.$$

The hazard function represents the instantaneous risk of experiencing the event of interest at time t given that an individual has survived until at least time t. For a small Δ, $h(t)\Delta$ can be interpreted as the probability of a subject failing in the next instant given he/she has survived until time t. In contrast to the survival function, which focuses on not experiencing the event, the hazard function focuses on experiencing the event of interest. Moreover, it should also be noted, unlike the survival function, that the hazard rate is not a probability.

The cumulative hazard function is another function that is also common in survival analysis. Like the hazard function, the cumulative hazard function is not a probability. The cumulative hazard function is defined as

$$H(t) = \int_0^t h(u)du,$$

which can be interpreted as the accumulated amount of hazard up to time t. Note that it is worth mentioning that there is an interesting relation between these three functions. In fact, given one of these functions, the other two can be uniquely determined. The relationships are: $S(t) = \exp\left\{-\int_0^t h(u)du\right\} = \exp\{-H(t)\}$ and $h(t) = -\frac{d}{dt}\log S(t) = \frac{f(t)}{S(t)}$.

1.1.2 Features of survival data

There are two special features of survival data that distinguish it from other types of data. The first feature is that survival times are usually positively skewed, and it is not reasonable to assume that data of this type have a normal distribution. Although transformation can be applied to make the distribution symmetric, it is appealing to consider other distributions that can fit naturally. In literature, there exist various distributions that have been commonly used to characterize survival time T. Some of these distributions are summarized in Table 1.1.

Table 1.1: Some commonly used parametric distributions

Distribution	$f(t)$	$S(t)$	$h(t)$
Exponential	$\lambda\exp(-\lambda t)$	$\exp(-\lambda t)$	λ
Weibull	$\rho\lambda t^{\rho-1}\exp(-\lambda t^{\rho})$	$\exp(-\lambda t^{\rho})$	$\rho\lambda t^{\rho-1}$
Lognormal	$\frac{1}{t\sigma\sqrt{2\pi}}\exp\left\{-\frac{(\log(t)-\mu)^2}{2\sigma^2}\right\}$	$1-\Phi(\frac{\log(t)-\mu}{\sigma})$	$\frac{f(t)}{S(t)}$

The other key feature of survival data is that exact event times of interest may not always be observed for all subjects under study and hence provide incomplete information. This phenomenon is known as censoring. In follow-up studies, various situations give rise to different kinds of censoring. In general, there are three main types of censoring: right censoring, left censoring, and interval censoring.

Right censoring: an observation is said to be right censored if the exact value of survival time is not known except that it is greater than a certain value. The three common conditions that yield right censored observation are: the study participant withdraws from the study, the study participant lost to follow-up during the study period, or the study participant has not experienced the event at the end of the study. There are three types of right censoring: Type I, Type II and random censoring. The latter is the most common case that occurs in medical trials, and hence in this work, we only consider random censoring. *Left censoring*: in contrast to the right censoring, this occurs when a subject experience an event of interest before the time of the first examination, and all that is known is the survival time is less than certain value. *Interval censoring*: refers to the case where we only have information that the event of interest occurred between two known time points,

but not the exact event time. This situation is a quite common phenomenon in many follow-up or longitudinal studies where subjects are assessed periodically for the response of interest. Note that interval censoring is a generalization of left and right censoring, with the lower limit of the interval is zero or the upper limit of the interval is infinity, respectively. Figure 1.1 show the three censoring type using a hypothetical data. From the figure, subject A experienced the event of interest after the end of the study so that only the solid part of the line is observed and the time is said to be right censored. For subject B, however, the event of interest has already occurred before visit time 4, so the period before the observed time (i.e., the dotted line) is not recorded and hence the survival time is said to be left censored at four weeks. Finally, subject C experienced the event between four and six weeks, so that the survival time is known to lie between four and six weeks instead of being observed exactly. Therefore, the survival time is said to be interval censored between four and six weeks.

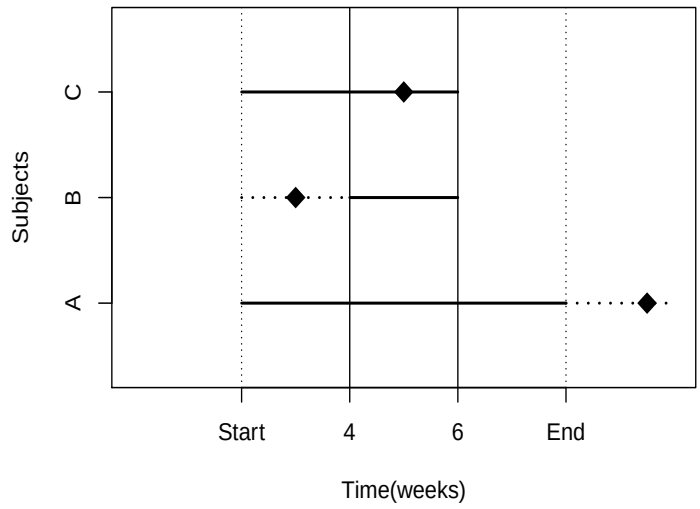

Figure 1.1: Plot to illustrate the three censoring mechanism; the solid diamonds denote the unobserved times when the event of interest occurred.

1.2 Survival estimation

In time-to-event data analysis, estimating quantities such as survival function is challenging due to censoring, a commonly encountered phenomenon, in survival

analysis, that may take different forms. The problem with censoring comes from the fact that the event time T is unknown for the censored subjects. The information we lose due to censoring has some effect on the estimation of survival quantities, and this may induce bias in the estimated quantities. The information we have about T and what can be done with it to make valid estimation and inference depends on the nature of censoring. The existing approaches that are proposed to deal with censored time-to-event data can be broadly classified into non-parametric, semi-parametric, and parametric methods. The difference between these three methods lies in the assumptions that we make regarding the distribution of the survival data. In this section, we give a review of some of the most commonly used estimation methods in survival analysis.

1.2.1 Nonparametric estimation

In the absence of censoring, the survival function can be estimated empirically as the proportion of individuals with event times greater than a given time t. This approach, however, is not valid when the data contain censored observations because the event times are not all observed. To this end, the Kaplan-Meier estimator (Kaplan and Meier; 1958), which is also well known as the product-limit estimator, is the most popular nonparametric method of estimating the survival function from the right censored data. This estimator is nonparametric, in the sense that no assumption for the survival distribution is needed. Kaplan and Meier (1958) derived this estimator using the method of maximum likelihood. In order to estimate the survival function at a given time t, let assume that a data set of size n is available. More specifically, the data consist of $\{(Y_i, \Delta_i), i = 1, 2, ..., n\}$ where $Y_i = \min(T_i, C_i)$ is the observed survival time, T_i is the time to the event of interest, C_i is the censoring time, $\Delta_i = I(T_i \leq C_i)$ is the censoring indicator. The Kaplan-Meier estimate of the survival function is given by

$$\hat{S}(t) = \prod_{i:\, Y_i \leq t} \left(1 - \frac{1}{\sum_{j=1}^{n} I(Y_j \geq Y_i)}\right)^{\Delta_i}. \tag{1.1}$$

This estimate is a step function with jumps at the observed event times. The properties of the Kaplan-Meier estimate of the survival function have been further explored by Breslow and Crowley (1974) and Meier (1975).

The main limitation of the Kaplan-Meier estimator is that it does not take the effect of the covariates into account. One possible solution is splitting the sample for each level of the covariate and apply the Kaplan-Meier estimator for each subsample. This approach is effective when the covariate is discrete but not recommended for the case of continuous covariates. To overcome this problem, in the last few decades, several generalizations to the Kaplan-Meier estimator have been proposed. Beran (1981) proposed a fully nonparametric estimator of the conditional survival function with right censored data. Given a vector of covariate X,

the Beran estimator which also known as a weighted Kaplan-Meier estimator is defined as

$$\hat{S}(t|x) = \prod_{i:\, Y_i \leq t} \left(1 - \frac{v_i(x)}{\sum_{j=1}^{n} v_j(x) I(Y_j \geq Y_i)}\right)^{\Delta_i}, \tag{1.2}$$

where, for a given bandwidth $b \equiv b_n$ and a given kernel density k, $v_i(x)$ are the most commonly used Nadaraya-Watson (Nadaraya (1965); Watson (1964)) weights

$$v_i(x) = \frac{k(\frac{X_i - x}{b})}{\sum_j k(\frac{X_j - x}{b})}.$$

The consistency and asymptotic normality of this estimator were established by several researchers (see, for example, Beran (1981); Dabrowska (1989); González-Manteiga and Cadarso-Suárez (1994); Van Keilegom and Veraverbeke (1997)).

The application of these methods, however, is limited to the case of right censored data. For example, if one is interested to apply these methods to interval censored data then he/she must have a way to compare and rank the observed intervals which is almost impossible. This is because it is not possible to rank two overlapping intervals. Therefore, to overcome this problem, the simplest solution is approximating the interval censored observations using the naïve approaches such as midpoint imputation that assume the event occurred at the midpoint of the observed interval. From this, we get exact failure time data rather than interval censored data, and it can be analyzed using the standard methods (for example, the Kaplan-Meier estimator) developed for right censored data. This approach is only reasonable when the periods between consecutive visits are short because its statistical properties strongly depend on the width of the interval between visits (Law and Brookmeyer; 1992). These authors noted that midpoint imputation, used to estimate the regression parameter in a proportional hazards model, might result in a biased estimate if the intervals are wide and varied. Also, this approach underestimates the standard error of the estimator because the midpoint imputation wrongly assumes that the failure times are exactly known (Kim; 2003). Turnbull (1976) proposed an extension of the Kaplan-Meier estimator to deal with different censoring types. The main idea of his new method is to establish the self-consistent equations and to solve them iteratively. Let $0 = s_0 < s_1 < s_2 < ... < s_m, m \leq 2n$, be the distinct increasing ordered values of $\{0, \{L_i\}_{i=1}^n, \{R_i\}_{i=1}^n\}$ where (L_i, R_i) denotes the observed event time interval of the i^{th} individual. Let p_j be the probability that the event of interest happened within $(s_{j-1}, s_j]$, $j = 1, ..., m$. The nonparametric maximum likelihood estimate of $\boldsymbol{p} = (p_1, ..., p_m)$ can be obtained by maximizing the likelihood function

$$L(\boldsymbol{p}; \alpha_{i,j}) = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{i,j} p_j,$$

where $\alpha_{i,j} = 1$ if $L_i < s_j \leq R_i$ and 0 otherwise. Several algorithms have been proposed to obtain the nonparametric maximum likelihood estimate of $\boldsymbol{p} = (p_1, ..., p_m)$ (see, for example, Turnbull (1976), Wellner and Zhan (1997), Wang (2008) and Anderson-Bergman (2017a)).

The nonparametric maximum likelihood estimate of the survival function is given by

$$\hat{S}(t) = \sum_{j:s_j > t} \hat{p}_j,$$

where $\hat{p}_j$ is the maximum likelihood estimate of p_j. This method, however, does not specify how the estimated probability is distributed within the Turnbull interval $(s_{j-1}, s_j]$.

In the presence of covariate(s), the Turnbull's estimator has the same difficulty as in the Kaplan-Meier estimator. Therefore, Dehghan and Duchesne (2011a) proposed a generalized method for estimating the conditional survival function of interval censored data. Without any mathematical details and using similar expression as in Dehghan and Duchesne (2011a), the maximum likelihood estimate of the conditional survival function is given by

$$\hat{S}(t|x) = \sum_{j:s_j > t} \hat{p}_j(x),$$

where $\hat{p}_j(x)$ is an estimate obtained by solving the self-consistency equation

$$p_j(x) = \sum_{i=1}^{n} v_i(x) \frac{\alpha_{i,j} p_j(x)}{\sum_{k=1}^{m} \alpha_{i,k} p_k(x)}, \ j = 1, ..., m,$$

where $v_i(x)$ are the Nadaraya-Watson weights defined above. One of the main limitations of this estimator is that it assigns positive probability to disjoint Turnbull intervals $(s_{j-1}, s_j]$ but how this probability is distributed within a Turnbull interval does not affect the likelihood and hence it is not necessarily unique. The other limitation is it requires selecting a bandwidth parameter needed for calculating the Nadaraya-Watson weights.

An alternative approach is a spline-based estimator proposed by Wu and Zhang (2012) and Wu et al. (2020). A spline is a function defined as a piecewise polynomial function in which the individual polynomials have the same degree. The piecewise polynomials are connected at the interval endpoints (knots) in such a way that a certain degree of smoothness of the resulting function is guaranteed. The spline is referred by its polynomial degree: a spline of degree one is linear spline, a spline of degree two is a quadratic spline, a spline of degree three is a cubic spline, and so forth. The type of polynomial and the number and position of knots is what defines

the type of spline. In the literature, there are numerous spline basis functions such as the B-spline basis function. The initial proposals of Wu and Zhang (2012) was the B-spline based sieve maximum likelihood estimator for a joint and marginal distribution functions. This method, however, involves complex constraints, making its numerical implementation notoriously difficult. Therefore, to simplify the computation, they proposed to compute the sieve maximum likelihood estimator using I-spline basis function.

Let the observed data for the i^{th} subject be $\{l_i, r_i, x_i, \delta_i^{(1)}, \delta_i^{(2)}, \delta_i^{(3)}, i = 1, 2, ..., n\}$ where l_i and r_i are the observed times, x_i is the observed value of the random variable X and $\delta_i^{(1)}, \delta_i^{(2)}$ and $\delta_i^{(3)}$ are the event indicator for left, interval and right censoring, respectively. Let I_j^a and M_j^a be I-spline and M-spline, respectively, where $M_j^a(t) = \frac{dI_j^a(t)}{dt}$. The definition of these spline basis functions are given in Wu and Zhang (2012). Wu and Zhang (2012) showed that $I_j^a(t) = \sum_{h=j+1}^{p_n+1} B_h^{a+1}(t)$, where B is the B-spline. Note that I_j^a is of degree a, both M_j^a and B_j^a are of degree $a-1$. The authors, using some algebra, showed that the joint cumulative distribution function of the random event time T and the random variable X, denoted by $F_n(.,.)$, and the marginal distribution function of X, denoted by $F_{n,2}(.)$, are given by

$$F_n(t,x) = \sum_{j=1}^{p_n-1} \sum_{k=1}^{q_n-1} \gamma_{j,k} I_j^{(1),a}(t) I_k^{(2),a}(x),$$

and

$$F_{n,2}(x) = \sum_{k=1}^{q_n-1} \Big\{ \sum_{j=1}^{p_n-1} (\gamma_{j,k} + \omega_k) \Big\} I_k^{(2),a}(x).$$

subject to the following constraints

$$\begin{aligned} &\gamma_{j,k} \geq 0 \text{ for } j = 1, 2, ..., p_n - 1, k = 1, 2, ..., q_n - 1, \\ &\omega_k \geq 0, k = 1, 2, ..., q_n - 1, \\ &\sum_{k=1}^{q_n-1} \sum_{j=1}^{p_n-1} \gamma_{j,k} + \sum_{k=1}^{q_n-1} \omega_k \leq 1. \end{aligned} \tag{1.3}$$

Given these constraints, the spline-based sieve maximum likelihood estimate can be obtained by maximizing the log-likelihood

$$\begin{aligned} l(\gamma, \omega) = \sum_{i=1}^{n} \Big[\delta_i^{(1)} \log \Big\{ \sum_{j=1}^{p_n-1} \sum_{k=1}^{q_n-1} \gamma_{j,k} I_j^{(1),a-1}(l_i) M_k^{(2),a-1}(x_i) \Big\} \\ + \delta_i^{(2)} \log \Big\{ \sum_{j=1}^{p_n-1} \sum_{k=1}^{q_n-1} \gamma_{j,k} I_j^{(1),a-1}(r_i) M_k^{(2),a-1}(x_i) \end{aligned}$$

$$- \sum_{j=1}^{p_n-1} \sum_{k=1}^{q_n-1} \gamma_{j,k} I_j^{(1),a-1}(l_i) M_k^{(2),a-1}(x_i) \Big\}$$
$$+ \delta_i^{(3)} \log \Big\{ \sum_{k=1}^{q_n-1} \Big\{ \sum_{j=1}^{p_n-1} (\gamma_{j,k} + \omega_k) \Big\} M_k^{(2),a}(x_i)$$
$$- \sum_{j=1}^{p_n-1} \sum_{k=1}^{q_n-1} \gamma_{j,k} I_j^{(1),a-1}(r_i) M_k^{(2),a-1}(x_i) \Big\} \Big], \tag{1.4}$$

where $\boldsymbol{\gamma} = \{\gamma_{j,k}\}_{j=1,2,...,p_n-1,k=1,2,...,q_n-1}, \boldsymbol{\omega} = \{\omega_k\}_{k=1,2,...,q_n-1}$. For the optimization, the authors suggested an efficient implementation method using the generalized gradient projection algorithm. From the resulting joint and marginal distribution function estimates, the conditional survival function estimate can be obtained using a numerical differentiation technique.

In this approach, like other spline-based methods, the problems of knots selection need to be addressed. More precisely, this is a problem of selecting the number and positions of the knots. To this end, for the event time component, the authors proposed choosing the spline knot sequence based on the observed times $\{(l_i, r_i), i = 1, ..., n\}$ with the number of interior knots be the closest integer to $n^{1/3}$. Specifically, let $O = \{l_i \delta_i^{(1)}, , \frac{l_i + r_i}{2} \delta_i^{(2)}, r_i \delta_i^{(3)}, i = 1, ., n\}$ meaning each member of O equals l_i for left censoring, $\frac{l_i + r_i}{2}$ for interval censoring and r_i for right censoring. From this, using a simple approach for knot placement, the authors suggested putting interior knots at the quantiles of O. Similarly, in the X direction, the knot sequence can be directly chosen based on the quantiles of $\{x_i, i = 1, ., n\}$.

1.2.2 Regression modeling in survival analysis

Regression modeling is the most commonly used method of relating the outcome of interest to covariates. This is important not only to determine the risk factors but it also plays a vital role in predicting the risk itself. To this end, the two well-known model classes in survival analysis are the Cox proportional hazards model and the accelerated failure time model.

The Cox proportional hazards model

The Cox proportional hazards model often called the Cox model, proposed by Cox in 1972 (Cox; 1972), is the most popular method which played a magnificent role for the advancement of both the theory and application of survival analysis for the past half century. For a given p-vector of covariates $\boldsymbol{X}$, the Cox model is written as

$$h(t|\boldsymbol{X}) = h_0(t) \exp(\boldsymbol{\beta}' \boldsymbol{X}), \tag{1.5}$$

where $h_0(.)$ is a baseline hazard function and $\boldsymbol{\beta}$ is a vector of regression coefficients. This model assumes a common baseline hazard for all subjects and it specifies the conditional hazard function as the product of the baseline hazard function and exponentially transformed covariates. Since no particular form of probability distribution is assumed for the survival times, the Cox model is a semiparametric model. In the Cox model, the time effect in the baseline hazard and the exponentiated covariate effect are separated, and hence the ratio of the hazard functions for two subjects with different covariates information is constant over time t. Let us consider two subjects with covariate measures $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$, then the ratio of hazard functions for these subjects is

$$\frac{h(t|\boldsymbol{X}_i)}{h(t|\boldsymbol{X}_j)} = \exp(\boldsymbol{\beta}'\boldsymbol{X}_i - \boldsymbol{\beta}'\boldsymbol{X}_j),$$

which is independent of time t. This assumption facilitates the interpretation of covariate effects. However, assessing this assumption is very important as it helps to check whether a fitted model adequately describes the data or not. This assumption can be checked using statistical tests and the scaled Schoenfeld residuals based diagnostics plots (Collett; 2015).

For right censored data, the regression coefficients in the Cox model, can be estimated using a partial likelihood maximization procedure introduced by Cox (1972). Assuming that there are no ties among event times and considering $h_0(t)$ as a nuisance parameter, Cox (1972) introduced the partial likelihood as

$$L(\boldsymbol{\beta}) = \prod_{i:\Delta_i=1} \frac{\exp(\boldsymbol{\beta}'\boldsymbol{X}_i)}{\sum_{k\in R(y_i)} \exp(\boldsymbol{\beta}'\boldsymbol{X}_k)},$$

where $R(y_i)$ is the number of observations who are still at risk to experience the event at time just prior to y_i. The regression parameter estimates can be obtained by maximizing the partial likelihood function, which is often done using the Newton-Raphson procedure. The asymptotic properties of the partial likelihood estimator have been extensively studied; see, for example, Tsiatis (1981) and Andersen and Gill (1982). For the case of interval censored data, several estimation algorithms have been proposed in the literature; see, for example, Finkelstein (1986), Pan (1999) and Anderson-Bergman (2017a).

An alternative method to the Cox model is the parametric proportional hazards model. The main difference between these two models is that the parametric proportional hazards model require a specific distributional assumption for the baseline hazard function. This means, for the parametric model, it is important to specify a particular shape for the baseline hazard function that is usually characterized by one or more parameters. Some of the commonly used distributions are the exponential and Weibull distribution. The exponential hazard assumes the hazard to

be constant over time, while the Weibull hazard assumes the hazard to increase or decrease over time; see Table 1.1. Even though the Cox model offers greater flexibility, if the assumed parametric model is correct, the parametric proportional hazards model will provide efficiency and the ability to extrapolate beyond the range of the event time T. The other difference is, unlike the semiparametric Cox model, the maximum likelihood estimation is used to estimate parameters of the parametric models. This is due to the fact that in the parametric model the baseline survival function is fully specified, which is not the case in the Cox model.

Once the conditional hazard function is given, using the relation $S(t) = \exp(-H(t))$, the corresponding condition survival function can be obtained as

$$S^{\mathcal{COX}}(t|\boldsymbol{X}) = S_0(t)^{\exp(\boldsymbol{\beta}'\boldsymbol{X})}, \tag{1.6}$$

where $S_0(t) = \exp\{-\int_0^t h_0(s)ds\}$ is the baseline survival function.

The accelerated failure time model

The proportional hazards model has been extensively used in many areas of application to investigate the relationship between covariates and the outcome of interest and to predict a risk. The proportional hazards assumption, however, is rather strong and may often be violated. In this situation, the accelerated failure time (AFT) model has been introduced as an important alternative to the proportional hazards model. The AFT model can be formulated using conditional survival function

$$S(t|\boldsymbol{X}) = S_0(t\exp(\boldsymbol{\beta}'\boldsymbol{X})),$$

where $\boldsymbol{\beta}$ is a vector of regression coefficients, $\boldsymbol{X}$ vectors of covariates, and $S_0(.)$ is a baseline survival function. This model shows that covariates affect survival functions by changing its time scale and hence the quantity $\exp(\boldsymbol{\beta}'\boldsymbol{X})$ is known as accelerating or decelerating factor.

An equivalent and popular formulation of AFT model is the log-linear form which is given by

$$\log(T) = \gamma_0 + \boldsymbol{\gamma}'\boldsymbol{X} + \sigma\epsilon, \tag{1.7}$$

where γ_0 is an intercept term, σ is a scale parameter, and ϵ is a random error term. Different choices for the error distribution yields a different model for survival time. For example, if the error distribution is the standard normal distribution, the survival time has a lognormal model. This model relates the logarithm of survival time linearly to a set of covariates, and thus it is appealing. The maximum likelihood estimates of the model parameters can be obtained using the numerical procedures such as the Newton-Raphson algorithms. In the context of the AFT model with unspecified error distribution, for both the right and interval-censored data, several estimation approaches, including an estimating equation approach

based on linear rank statistics, have been proposed; see, for example, Kalbfleisch and Prentice (2011); Sun (2007).

1.3 Survival estimation in the presence of cure fraction

In the classical methods of survival data analysis, for example, the proportional hazards model and the AFT model, one of the most important assumptions is that all subjects in the study will eventually experience the event of interest if the follow-up period is long enough. This means

$$\lim_{t\to\infty} S(t) = 0.$$

With the recent advancement in medical technology, this assumption, however, might not be appropriate in many situations. This is, for instance, the case of cancer research when the main endpoint is time to relapse. In this scenario, a fraction of patients can be cured and hence will never relapse. In this situation, the classical survival models may not be adequate to fit survival data with cure fraction. Alternatively, several models widely called *cure models* have been proposed to accommodate a cured fraction. In the literature, there are two major model classes to fit survival data with cure fraction: the mixture cure rate model and the promotion time cure model. In the upcoming subsections, we present a short review of these models.

1.3.1 Mixture cure model

The mixture cure model (MCM) was first introduced by Boag (1949) to study mouth cancer. In this model, it is assumed that the population consists of two types of subjects: uncured (susceptible) subjects who will experience the event of interest, and cured (non-susceptible) subjects who will never experience it. The mixture cure model is a combination of two models, one for the probability of cure sometimes called "the incidence model" and one for the event time of uncured subjects sometimes called "the latency model". Let B be the cure indicator with $B = 0$ when the subject is cured and $B = 1$ when the subject is uncured. Let $p = P(B = 1)$ be the probability of being uncured, and the probability of being cured be $P(B = 0) = 1-p$. The mixture cure model basically allows to investigate the effect of the covariates both on the cure probability and on the survival function of uncured subjects. The survival function in the mixture cure model is given by

$$S^{\mathcal{MCM}}(t|\boldsymbol{X}, \boldsymbol{Z}) = p(\boldsymbol{Z})S_u(t|\boldsymbol{X}) + (1 - p(\boldsymbol{Z})), \tag{1.8}$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the two vectors of covariates that are either identical or partly or entirely different from each other, $S^{\mathcal{MCM}}(t|\boldsymbol{X}, \boldsymbol{Z}) = P(T > t|\boldsymbol{X}, \boldsymbol{Z})$ is the conditional population survival function and $p(\boldsymbol{Z}) = \mathrm{P}(B = 1|\boldsymbol{Z})$ is the conditional

probability of being uncured (the incidence). The function $S_u(t|\boldsymbol{X}) = P(T > t|\boldsymbol{X}, B = 1)$ is the conditional survival function of the susceptible subjects (the latency). It is clear that $\lim_{t\to\infty} S^{\mathcal{MCM}}(t|\boldsymbol{X}, \boldsymbol{Z}) = 1 - p(\boldsymbol{Z})$. Furthermore, the mixture cure model reduces to the classical survival model if all the study subjects are susceptible.

The mixture cure model involves separately modeling the cure probability and the survival distribution of the uncured patients. In literature, for both incidence and latency model, parametric, semiparametric and nonparametric methods have been proposed by several authors. The incidence part of the mixture cure model is usually modeled with logistic regression, which can be given by

$$p(\boldsymbol{Z}) = \frac{\exp(\gamma_0 + \boldsymbol{\gamma}'\boldsymbol{Z})}{1 + \exp(\gamma_0 + \boldsymbol{\gamma}'\boldsymbol{Z})},$$

where γ_0 is an intercept term. As we stated before, other models, such as the flexible nonparametric single index model proposed by Amico et al. (2019), can also be used for the incidence part of the mixture cure model. The latency part of the mixture cure model can be modeled using either the most popular proportional hazards model

$$S_u(t|\boldsymbol{X}) = S_0(t)^{\exp(\boldsymbol{\beta}'\boldsymbol{X})},$$

or the accelerated failure time model

$$S_u(t|\boldsymbol{X}) = S_0(t\exp(\boldsymbol{\beta}'\boldsymbol{X})).$$

As for the estimation of the unknown model parameters, including the popular Expectation-Maximization (EM), various algorithms have been proposed in the literature. A comprehensive review of the current estimation approaches and inference procedures of this model is given by Amico and Van Keilegom (2018b).

1.3.2 Promotion time cure model

Yakovlev et al. (1993) developed an alternative to the mixture cure model called promotion time model also known as "bounded cumulative hazard model" or "proportional hazard cure model" (BHM). This model is an extension of the classical Cox model (Cox; 1972) designed to handle the presence of cure fraction. The promotion time model has been extensively investigated by, among others, Yakovlev et al. (1996); Tsodikov (1998a,b); Chen et al. (1999); Tsodikov (2001). It is interesting that this model has both mathematical and biological derivations. As presented in Tsodikov (1998b, 2001), the mathematical derivation was based on the property of the cumulative hazard function $H(t)$. Unlike the cumulative hazard function of survival data without cure fraction, the limiting value of $H(t)$ is finite when there are cured subjects in the data. This means $\lim_{t\to\infty} H(t) = \theta < \infty$, indicating that the cumulative hazard is bounded from above by θ. In order to take

the boundedness into account, it is convenient to write the cumulative hazard function as $H(t) = \theta F(t)$, where $F(t)$ is a proper cumulative distribution function (i.e., $\lim_{t\to\infty} F(t) = 1$). This lead to an improper survival function that can be obtained as $S(t) = \exp(-H(t)) = \exp(-\theta F(t))$. The biological derivation of this model was provided by Yakovlev et al. (1996) and Chen et al. (1999). The formulation was based on the assumption that the number $N \geq 0$ of cancer cells remaining after treatment to follow a Poisson distribution with parameter θ (Chen et al.; 1999). Let denote the promotion time for the i^{th} cell by Y_i, which is the time to produce a detectable cancer cell mass. Assuming that the random variables $Y_i, i = 1, 2, ...N$ are independent and identically distributed with a cumulative distribution $F(t)$ and are independent of N, the survival function of the time to relapse of cancer that defined as $T = \min\{Y_i, 0 \leq i \leq N\}$ with $P(Y_0 = \infty) = 1$ is given by

$$\begin{aligned} S^{\mathcal{BHM}}(t) &= P(N = 0) + P(Y_1 \geq t, ..., Y_N \geq t, N \geq 1) \\ &= P(N = 0) + \sum_{k=1}^{\infty} P(Y_1 \geq t, ..., Y_k \geq t) P(N = k) \\ &= \exp(-\theta) + \sum_{k=1}^{\infty} \{1 - F(t)\}^k \frac{\exp(-\theta)\theta^k}{k!} \\ &= \exp(-\theta + \theta S(t)) = \exp(-\theta F(t)). \end{aligned}$$

Incorporating the covariate(s) effects, this can be written as

$$S^{\mathcal{BHM}}(t|\boldsymbol{X}) = \exp\{-\theta(\boldsymbol{X})F(t)\}, \tag{1.9}$$

where $\theta(\boldsymbol{X})$ is a positive function which captures the effect of the covariates $\boldsymbol{X}$ on the population survival function and is usually expressed as $\exp(\boldsymbol{X}'\boldsymbol{\beta})$. Using the equation above, the cure fraction can be derived as $P(B = 0|\boldsymbol{X}) = \lim_{t\to\infty} S^{\mathcal{BHM}}(t|\boldsymbol{X}) = \exp\{-\theta(\boldsymbol{X})\}$. The advantages of this class of cure models, in comparison to the mixture cure model, is that it has a proportional hazard structure since the covariates are involved only through the cure rate parameter, and thus has an appealing interpretation. Furthermore, this class of model is derived from a natural biological motivation and it has a simple structure for survival function which makes the computations attractive (Chen et al.; 1999). Its limitation, however, is that the separate effect of the covariates on the probability of being cured and on the event time of susceptibles cannot be estimated. For a detailed review about both the classical and Bayesian estimation methods of this model, the reader is referred to Amico and Van Keilegom (2018b).

1.4 Evaluating prediction accuracy

Survival models are regression models that are useful to understand the associations between a set of baseline factors to a given event of interest. It allows us

to predict the risk of developing a future event, given the baseline factors. In clinical medicine, for example, this is important to develop appropriate intervention strategies for those who are at high risk of an adverse health outcome. Like many other statistical models, however, ensuring their validity is crucial before using them for clinical decision making. Assessing the validity of survival models often involves measuring the predictive accuracy, the ability of a model to predict the risk of a future event given a set of baseline factors. In literature, there are several ways for assessing the predictive accuracy of survival models which often described by two components: calibration and discrimination. Discrimination characterizes how well the model can classify subjects into one of the two groups (e.g., diseased and non-diseased), whereas calibration measures the degree of closeness between the predicted probabilities and the observed outcome. There are a variety of methods to evaluate the discriminatory ability of a survival model. The most commonly used are the receiver operating characteristic curve and the associated summary measures. In this thesis, we propose different receiver operating characteristic curves and their corresponding summary measures estimation methods for censored time-to-event data.

1.4.1 The receiver operating characteristic (ROC) curve

The ROC curve is the most popular and well-developed statistical tool to describe the predictive ability of a continuous marker. The ROC curve was first introduced during World War II to assess the radar receiver operators ability to correctly detect incoming enemy aircraft, and later it was used in signal detection theory (Green et al.; 1966; Egan; 1975). Since then, the ROC has become the most widely used and accepted tool in various application areas. The first application of ROC methodology in medical diagnostic tests date back to the beginning of the 1960s (Lusted; 1960). However, it was in the early 1980s, after publication of the text by (Swets and Pickett; 1982), that they became very popular, especially in the field of radiology (Metz; 1989). Some other application areas includes psychology (Swets; 1973), epidemiology (Aoki et al.; 1997), biomedical informatics (Lasko et al.; 2005) and machine learning (Spackman; 1989). Nowadays, the use of ROC curves to assess the ability of continuous diagnostic markers to identify individuals with and without the disease of interest has been increased (Shapiro; 1999; Greiner et al.; 2000; Zhou and Harezlak; 2002; Pepe; 2003). Without loss of generality, throughout this thesis, we use terminologies and examples from medical applications.

The classical ROC analysis was designed for assessing the accuracy of a continuous diagnostic test or marker in predicting a binary outcome (e.g., presence or absence of disease). Let D denote a disease status with $D = 1$ indicating presence (cases) and $D = 0$ indicating absence of disease (controls), and M denote a continuous marker with the conventional assumption that higher value of M is indicative of the presence of disease. For a given threshold $m \in \mathbb{R}$, the true pos-

itive rate (TPR) or sensitivity is the probability that the marker value is greater than the threshold given the disease is present and the false positive rate (FPR) or one minus specificity is the probability that the marker value is greater than the threshold given the disease is absent. Mathematically, these are written as

$$\begin{aligned} \text{TPR}(m) &= P(M > m|D = 1), \\ \text{FPR}(m) &= P(M > m|D = 0). \end{aligned} \tag{1.10}$$

From these definitions, both TPR and FPR are not dependent on disease prevalence because the TPR and FPR calculations involve subjects with and without the disease, respectively. This property is important since it allows transferring results from one study to another population with a different disease prevalence rate. The ROC curve and the related summary measures also inherit this property and hence are prevalence independent.

The ROC curve is defined as a plot of TPR versus FPR for all possible classification threshold values. Equivalently, in terms of function, it can be defined as

$$\text{ROC}(u) = \text{TPR}\{\text{FPR}^{-1}(u)\}, \tag{1.11}$$

where $u \in [0, 1]$ and $\text{FPR}^{-1}(u) = m$ is a threshold value such that $u = \text{FPR}(m)$.

The ROC curve depicts the trade-offs between sensitivity and specificity of the marker with regards to the cutoff points. The ROC analysis is useful to evaluate the discriminatory ability of a continuous marker, obtain an optimal cutoff point, and compare the relative performance of two (or more) diagnostic markers. It also generally useful as a graphical characterization of the magnitude of separation of marker distributions in the case and control groups. Theoretically, the ROC curve of a perfect diagnostic marker would be the one that rises steeply along the left axis to the point (FPR = 0, TPR = 1), while an uninformative marker has an ROC curve that is the diagonal line (i.e., the TPR is equal to the FPR for all cutoff values). The former indicates that the measurements of case and control have no overlap, whereas the latter shows that the two groups have identical distributions. In practical applications, the ROC curves typically fall somewhere in between these two theoretical cases. A graphical example illustrating these situations is given in Figure 1.2.

The definitions given above are specific to a cross-sectional data type with binary outcomes. In practice, for example in time-to-event studies, we often encounter disease status that changes with time. Therefore, it is necessary to incorporate the time dimension in the definitions of prediction accuracy measures. In the past two decades, various time-dependent TPR and FPR definitions have been proposed in the literature. This led to several time-dependent ROC curves and time-dependent summary measures definitions. In this section, we review the well-known extensions of the TPR and FPR definitions for survival outcomes.

Cumulative TPR and *Dynamic* FPR

Heagerty et al. (2000) proposed the most popular extensions of TPR and FPR definitions to the time-to-event outcomes. Their idea is dichotomizing the outcome at any fixed time of interest, t, and define *cumulative* cases (C) as subjects who experience the event before time t and *dynamic* controls (D) as those who remain event-free through time t. For a given cutoff point m, the *cumulative* TPR (TPR_t) and the *dynamic* FPR (FPR_t) at fixed time t are given by

$$\begin{aligned} \text{TPR}_t(m) &= P(M > m | T \leq t), \\ \text{FPR}_t(m) &= P(M > m | T > t). \end{aligned} \tag{1.12}$$

Under these definitions, the time-dependent *cumulative/dynamic* ROC (ROC_t) is defined as

$$\text{ROC}_t(u) = \text{TPR}_t\{[\text{FPR}_t]^{-1}(u)\}. \tag{1.13}$$

These definitions are more appropriate when the scientific interest lies in discriminating between individuals who experience the event prior to a given time t and those that are event-free beyond time t. This type of discrimination has more clinical relevance than the other definitions (Blanche et al.; 2013). The *cumulative/-dynamic* definition, however, uses redundant information because each individual plays the role of a control at an earlier time and then contributes as a case for later times (Heagerty and Zheng; 2005; Kamarudin et al.; 2017). When the event times, T, are observed for all the study subjects, the time-dependent TPR, FPR, and the corresponding ROC can be easily estimated using the empirical proportions. In the presence of censoring, however, estimating these quantities is challenging. The problem with censoring comes from the fact that T is unknown for the censored subjects. In the literature, there are several approaches for estimating *cumulative* TPR and *dynamic* FPR and the corresponding time-dependent ROC curve that take censoring into account. For comprehensive review of the literature, see Kamarudin et al. (2017) and Blanche et al. (2013). In Section 2.1, we offer a review of some of the existing proposals of the time-dependent *cumulative/dynamic* ROC curve estimation methods.

Incident TPR and *Dynamic* FPR

The *incident* TPR and *dynamic* FPR are defined by dichotomizing the risk set at time t into those who experience the event (cases) and those who are event-free (controls). In this approach, subjects with $T = t$ are considered as *incident* cases (I), whereas subjects with $T > t$ are considered as *dynamic* controls (D). Using these, for a given cutoff point m, the time-dependent definition for the *incident* TRP (TPR_t^I) is

$$\text{TPR}_t^I(m) = P(M > m | T = t), \tag{1.14}$$

and the resulting time-dependent ROC curve ($\text{ROC}_t^{I/D}$) can be written as

$$\text{ROC}_t^{I/D}(u) = \text{TPR}_t^I\{[\text{FPR}_t^D]^{-1}(u)\}. \tag{1.15}$$

As mentioned before, these definitions are based on classification of the risk set at time t into case when $T = t$ and control when $T > t$, and hence it is a natural companion to hazard models (Heagerty and Zheng; 2005). The *I/D* ROC definition is useful when the interest is distinguishing subjects failing at t and those surviving after this time. In addition, these definitions allow an extension to time-dependent or longitudinal marker. In this approach, however, a subject can play the role of a control for an early time, $T > t$ and then play the role of case when $T = t$. Also, a subject who experience the event prior to a given time t can not play any role. Several authors have proposed estimators of the time-dependent ROC curve relying on the *incident* definition of cases and *dynamic* definition of controls, see, for example, Martínez-Camblor and Pardo-Fernández (2018); Shen et al. (2015); Saha-Chaudhuri and Heagerty (2013); Heagerty and Zheng (2005).

Incident TPR and *Static* FPR

In the *incident* TPR and *static* FPR definition each subject does not change event status and is treated as either a case or a control. Here, *incident* cases (I) are classified according to the time at which the event occurs, t, and *static* controls (S) are defined as those subjects who are event-free through a fixed time t^*. The rationale of using the fixed follow-up is because the end point t^* is pre-specified and it is considered a long enough time to eliminate patients with preclinical diseases from the control group. Given a threshold m, the time-dependent *static* FPR ($\text{FPR}_{t^*}^S$) is defined as

$$\text{FPR}_{t^*}^S(m) = P(M > m | T > t^*), \tag{1.16}$$

and the corresponding ROC curve ($\text{ROC}_t^{I/S}$) can be written as

$$\text{ROC}_t^{I/S}(u) = \text{TPR}_t^I\{[\text{FPR}_{t^*}^S]^{-1}(u)\}. \tag{1.17}$$

These definitions facilitate the use of standard regression approaches for describing TPR and FPR because the event time, T, can simply be used as a covariate (Heagerty and Zheng; 2005). Furthermore, it allow extension to time-dependent or longitudinal marker. Several approaches have been proposed for estimation, see, for example, Pepe et al. (2008); Cai et al. (2006); Zheng and Heagerty (2004); Slate and Turnbull (2000); Etzioni et al. (1999).

1.4.2 Some ROC curve summary measures

The ROC analysis gives numerous statistical approaches to summarize the prediction accuracy of a marker. Summarizing the ROC curve with a single numeric

value has some important uses, such as providing a concise way of communicating the diagnostic performance, facilitating comparisons, and serving as the basis of inferential statistics for comparing ROC curves (Pepe; 2003). Several ROC curve based summary indices have been proposed in the literature to evaluate the overall predictive performance. In this thesis, we focus on the two most commonly used ROC curve summary indices: the area under the ROC curve and the Youden index. This section provides a brief overview of each of these indices.

Area under the ROC curve (AUC)

The AUC is the most widely used global measure of the overall ability of the diagnostic marker to distinguish among individuals with and without the event of interest. It is defined as

$$\text{AUC} = \int_0^1 \text{ROC}(u)du.$$

The AUC value ranges between 0.5 and 1 and marker with higher AUC value would be preferable. For a marker with perfect discrimination between subjects with and without the event of interest, the theoretical AUC will be 1. A value of 0.5 for AUC indicates that the ROC curve will fall on the diagonal line and hence suggests that the diagnostic marker has no discriminatory ability (see Figure 1.2).

The AUC has some interesting interpretations. Most commonly, assuming that a higher marker value leads to a positive diagnosis, it is interpreted as the probability that a randomly selected subject with the event has larger marker value than a randomly chosen subject without the event (Hanley and McNeil; 1982). Alternatively, AUC can be interpreted as the average value of sensitivity for all possible values of specificity or the average of specificity for all possible values of sensitivity (Pepe (2003), Zhou et al. (2011)). Although the former interpretation of the AUC is interesting, its clinical relevance is under some critics and hence it is suggested to use the latter one (see, for example, Pepe (2003)). Dodd (2001) suggested to calculate a weighted average, weighting certain parts of the false positive rate domain more than others.

The time-dependent true positive and false positive rates together with the corresponding time-dependent ROC functions are defined above. From these, the time-dependent AUC is given by

$$\text{AUC}_t = \int_0^1 \text{ROC}_t(u)du,$$

where $\text{ROC}_t(.)$ is any of the previously defined time-dependent ROC functions. The time-dependent AUC can be straightforwardly estimated using either the numerical integration methods such as the trapezoid rule or the concordance based formula (Li et al.; 2018).

Youden index (J)

The Youden index is another most commonly used metric to evaluate the overall classification accuracy of a marker. It was first introduced by Youden (1950). The Youden index is defined as the maximum vertical distance between the ROC curve and the diagonal line (Pepe; 2003). Formally, it is defined as

$$\mathrm{J} = \max_{m}\{\mathrm{TPR}(m) - \mathrm{FPR}(m)\}.$$

It indicates how far the ROC curve is from the uninformative curve. It ranges from 0 (completely uninformative marker) to 1 (perfect marker); see Figure 1.2. The Youden index is also linked to the area of the shaded region (see Figure 1.3) associated with the ROC; see Liu (2012) and Hilden and Glasziou (1996). From the figure, the area of shaded region is the sum of area of the shaded triangles above and below the diagonal line. Using simple geometry, this area equals to $(J + 1)/2$ which is 0.5 for uninformative and 1 for perfect test. Therefore, maximizing the Youden function is equivalent to maximizing this area. Other interpretations can be found in the literature; see, for example, Hilden and Glasziou (1996). The Youden index has an attractive feature not present in the AUC. It provides the point in the ROC curve with the maximum vertical distance to the diagonal (informative) line. The marker value associated with this point is typically used as the optimal cutoff value, i.e., the marker value that maximizes the difference between the true and the false positives rates. The clinicians use this cutoff in order to classify subjects according to the marker value (less than cutoff = healthy, greater than cutoff = diseased). Note that many other criteria for summarizing the ROC curve and for obtaining the "optimal" cutoff value are available in the literature; see, for example, Sima and Gönen (2013), Liu and Jin (2015) and Rota et al. (2015).

The time-dependent Youden index, which corresponds to the time-dependent TPR and FPR defined above, can be given by

$$\mathrm{J}_t = \max_{m}\{\mathrm{TPR}_t(m) - \mathrm{FPR}_t(m)\},$$

where $\mathrm{TPR}_t(.)$ and $\mathrm{FPR}_t(.)$ are either C/D, I/D or I/S combination of the time-dependent TPR and FPR definitions. The most commonly used method of calculating this index is by maximizing the difference between time-dependent TPR and FPR empirically. For right censored time-to-event data, Rota et al. (2015) compared this method with other methods using simulations.

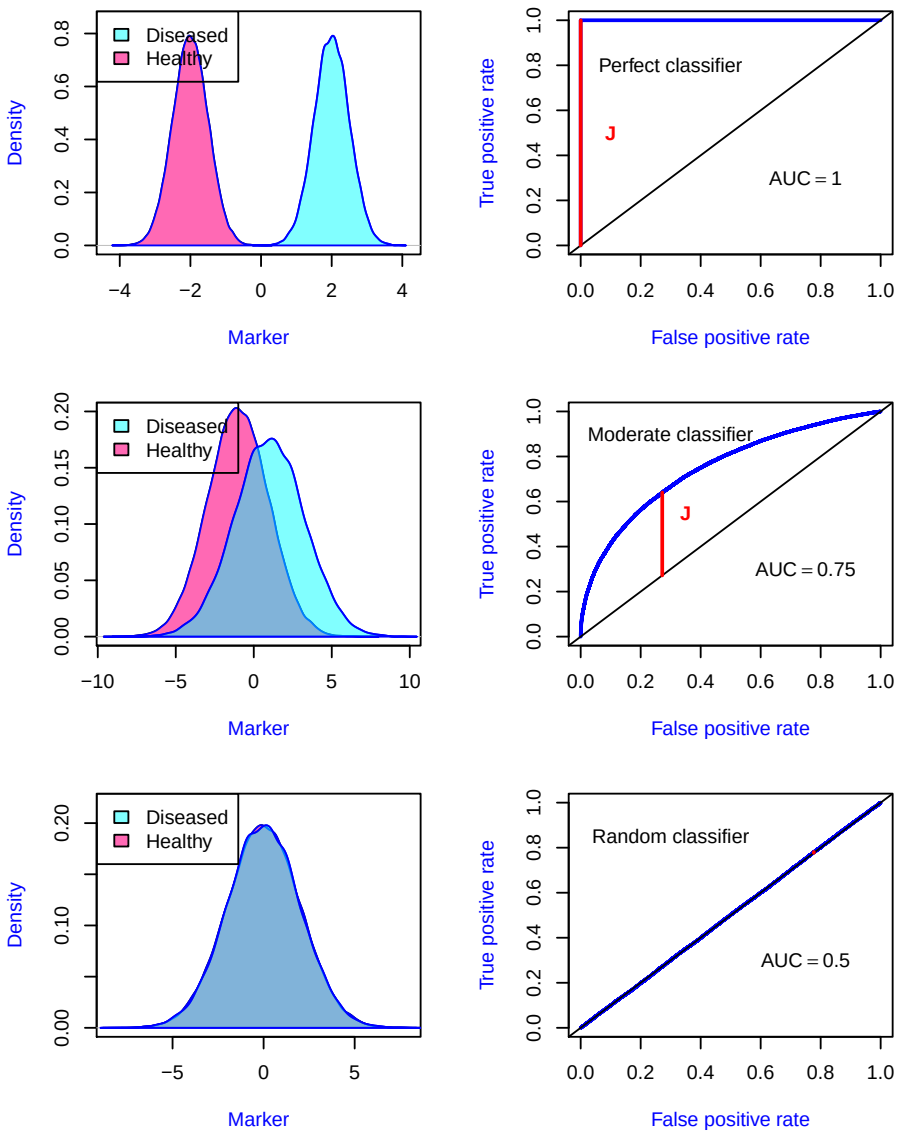

Figure 1.2: Hypothetical marker distributions in healthy and diseased populations and the corresponding ROC curves.

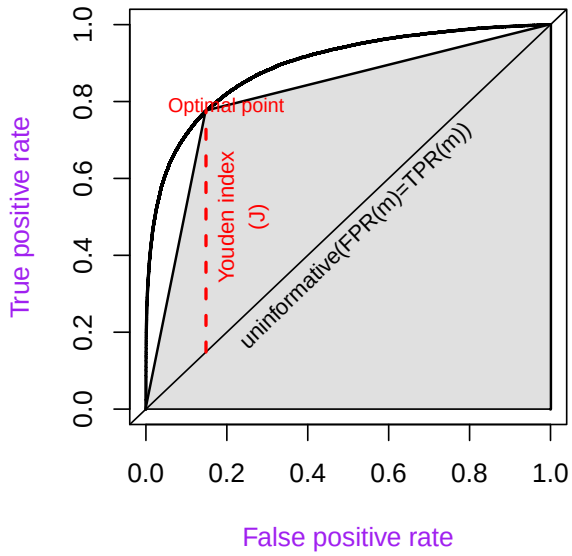

Figure 1.3: A ROC curve with the Youden index and the corresponding cutoff value.

1.5 The bootstrap approach

Statistical inference involves using data from a sample to draw appropriate conclusions about a population, which primarily depends on having some idea about the variability of the sample estimates. To this end, one can use either an asymptotic theory or the well-known bootstrap approach. However, in many situations, including, for example, the time-dependent AUC estimator, obtaining the asymptotic distribution or at least a simple closed-form solution for the asymptotic variance is very difficult or almost impossible. To alleviate this problem, Efron (1979) introduced a modern alternative to the classical approach called the bootstrap method. The term 'bootstrap' originates from the phrase "pulling oneself up by one's bootstraps" means approximate, for example, the sampling distribution and standard error of an estimator from the sample data at hand only (i.e., without using any theory). The original intent of this method was to estimate the distribution of a given sample statistic, then developed to use it for standard error estimation and to make statistical inferences. The basic idea of the bootstrap method is to generate a large number of pseudo-samples known as bootstrap samples by randomly drawing n observations with replacement from the sample data at hand. These

pseudo-samples are useful to get the sampling distribution of an estimator or test statistic and then obtain estimates of standard error or confidence intervals and conduct hypothesis tests based on these results.

1.5.1 Bootstrapping

There are mainly two forms of bootstrapping: parametric and nonparametric bootstrapping. The former assumes that the data comes from a known population distribution with unknown parameters, whereas the latter does not make any assumption about the population distribution. In this thesis, we use the nonparametric bootstrapping since it is more general and makes fewer assumptions. To describe the bootstrapping procedure, let us assume that the observed data set consists of $\{X_i, i = 1, 2..., n\}$, and suppose we are interested in some statistic $\hat{\theta}$ as an estimate of the corresponding population parameter θ. The algorithm for obtaining the bootstrap estimate of the variability of a given statistic, for example, the standard deviation of $\hat{\theta}$, is as follows

- Draw a large number, say B, of random samples of size n with replacement from the original sample $\{X_i, i = 1, 2..., n\}$.
- For each bootstrap samples, estimate the parameter θ to get $\{\hat{\theta}_b, b = 1, \cdots, B\}$.
- The empirical standard deviation of the bootstrap estimates is calculated using

$$S_B = \sqrt{B^{-1} \sum_b \left(\hat{\theta}_b - B^{-1} \sum_b \hat{\theta}_b\right)^2}.$$

  This can be used as an important estimator of the actual standard deviation of $\hat{\theta}$.

To give an intuitive explanation of why the bootstrap works, let us assume that the X_is are i.i.d. with a distribution function F and let G be the distribution of $\hat{\theta}$ based on sampling from F. Since F is unknown in practice, if we want to estimate G, then it would be natural to sample from the distribution that we do have available, which is the empirical distribution function $\hat{F}$. By taking a large sample from $\hat{F}$, we can well approximate G with $\hat{G}$. Since we can choose B as large as we like, we can ignore the approximation error and assume that G is known. Thus we can use this to obtain an estimator of the quantiles and confidence intervals. The question here, however, is how many bootstrap samples we need to get an acceptable variance estimate. This problem is important since both the accuracy of the variance estimator and the resources need to carry out the bootstrapping increases with an increase in the number of bootstrap samples. As for the variance estimation, Efron (1981) recommended that a large number of bootstrap samples would be 200, whereas the number of bootstrap samples he suggested for approximating confidence intervals is 1000 or more.

1.5.2 Bootstrap confidence intervals

In the literature, there exist various methods to construct bootstrap confidence intervals using the standard error estimate given above. This thesis presents two of the most common approaches for estimating bootstrap confidence intervals: the standard (or normal-theory) and the percentile.

The standard method for constructing a bootstrap confidence interval assumes that the sampling distribution of $\hat{\theta}$ is approximately normal. Using bootstrap estimates of the sample standard error for $\hat{\theta}$, the $(1-\alpha)100\%$ confidence interval of the population parameter θ have the following form

$$\hat{\theta} \pm z_{1-\alpha/2} S_B,$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution.

The simplest method of constructing the bootstrap confidence intervals is the percentile method, which is based on the empirical distribution of the bootstrap estimates. This method does not require to compute the standard error of $\hat{\theta}$. The percentile method is fully nonparametric because it does not make any specific distributional assumption for the estimator. The $(1-\alpha)100\%$ bootstrap percentile confidence interval for the population parameter θ is given by

$$[\hat{\theta}_B(\alpha/2), \hat{\theta}_B(1-\alpha/2)],$$

where $\hat{\theta}_B(\alpha)$ is the $\alpha-$quantile of the B bootstrap sample estimates, which means the $(B.\alpha)^{\text{th}}$ value of the ordered $\{\hat{\theta}_b, b = 1, \cdots, B\}$. Finally, it should be noted that the above percentile confidence interval will have the claimed coverage probability if there exists a monotonic transformation function $g(.)$ such that $g(\hat{\theta}) - g(\theta)$ and $g(\hat{\theta}_B) - g(\hat{\theta})$ have the same distribution that is symmetric about the origin (Efron; 1981; Diciccio and Romano; 1988).

1.6 Outline of the thesis

The main aim of this thesis is to propose and investigate time-dependent ROC curve and the corresponding summary measures estimators for censored time-to-event data. The contributions in this thesis are based on the *cumulative* TPR and *dynamic* FPR definitions. The contents of this thesis are organized as follows.

In Chapter 2, we address issues in measuring the classification accuracy of right censored time-to-event data with cure fraction. The objective of this chapter is twofold. The first objective is to investigate the validity of the time-dependent AUC estimation method proposed for the classical right censored data in the presence of cure fraction. To this end, we derive definitions for time-dependent TPR

and FPR in the presence of cure fraction and show that it is the same as the classical one proposed by Heagerty et al. (2000). For investigation purposes, we consider the simple nonparametric time-dependent AUC estimation method that was proposed by Li et al. (2018). The second objective is, motivated by the current practice of deriving risk tools from survival models that do not acknowledge the presence of cure, to examine the loss, in terms of discriminatory performance, of using a classical survival model to derive risk tools. This chapter is based on Beyene et al. (2019).

The third chapter presents a new time-dependent ROC curve estimation methods for right censored time-to-event data. The objective is to propose a smooth time-dependent ROC estimator. To this end, we first derive an empirical ROC estimator using the idea of imputation introduced by Li et al. (2018) and Martínez-Camblor et al. (2016). Then, relying on the well-known kernel techniques used to estimate cumulative distribution functions, we propose a new nonparametric smooth ROC estimator based on a weighted kernel smoothing. To further refine the smooth ROC estimator, we will consider a boundary correction based on a quantile transformation method. We also derive some asymptotic properties of the proposed estimators and propose a bootstrap procedure for estimating the standard deviation and constructing confidence intervals. For the parameter needed to smooth the ROC, we propose data-driven bandwidth selection methods. The proposed methods are implemented in R package `cenROC` (Beyene and El Ghouch; 2020a). This chapter discusses the methods presented in Beyene and El Ghouch (2020b).

Methods for evaluating the predictive accuracy of models for interval censored time-to-event data, relative to the right censored data, are scarce in the statistical literature. To address this issue, in Chapter 4, we propose some new time-dependent ROC curves and the corresponding summary index estimators for interval censored survival data. In the proposed method, taking the interval censored nature of the data into account, only unknown event status are imputed using a weight function that depends on a conditional survival function that needs to be estimated. To this end, in this chapter, we propose both a model-based and a nonparametric sieve maximum likelihood approach to estimate a conditional survival function. The ROC estimators are both empirical (non-smoothed) and smoothed as in Chapter 3. Furthermore, the time-dependent AUC and the time-dependent Youden index are the two indices employed to summarize the ROC curve. To estimate the variance and make inference about the ROC curve and the associated summary measures, we propose a bootstrap approach. We have implemented these methods in R package `cenROC` (Beyene and El Ghouch; 2020a). This chapter presents the results presented in Beyene and El Ghouch (2020c).

This thesis concludes with Chapter 5, which summarizes the main findings and provides some directions for possible future research works.

CHAPTER 2

On the Validity of Time-Dependent AUC Estimation in the Presence of Cure Fraction

Abstract

During the last decades, several approaches have been proposed to estimate the time-dependent area under the receiver operating characteristic curve (AUC) of risk tools derived from survival data. The validity of these estimators relies on some regularity assumptions among which a survival function being proper. In practice, this assumption is not always satisfied because a fraction of the population may not be susceptible to experience the event of interest even for long follow-up. Studying the sensitivity of the proposed estimators to the violation of this assumption is of substantial interest. In this chapter, we investigate the performance of a nonparametric simple estimator, developed for classical survival data, in the case when the population exhibits a cure fraction. Motivated from the current practice of deriving risk tools in oncology and cardiovascular disease prevention, we also assess the loss, in terms of predictive performance, when deriving risk tools from survival models that do not acknowledge the presence of cure. The simulation results show that the investigated method is valid even under the presence of cure. They also show that risk tools derived from survival models that ignore the presence of cure have smaller AUC compared to those derived from survival models that acknowledge the presence of cure. This was also attested with a real data analysis from a breast cancer study.

This chapter is based on Beyene KM, El Ghouch A, Oulhaj A. (2019). On the validity of time-dependent AUC estimation in the presence of cure fraction. *Biometrical Journal*, **61**(6): 1430-1447.

2.1 Introduction

Prognostic tools derived from survival models are becoming more commonly used in many areas. In healthcare applications, predicting patient's risk of developing a disease of interest within a pre-specified time horizon is usually of major interest. It assists in decision making such as optimizing clinical management of patients, defining better strategies for treatments, assessing cost effectiveness of potential therapies from health economic point of view and assisting in recruitment of patients, in clinical trials, by identifying subjects with higher risk of developing the outcome of interest. Due to the increasing role of risk predictions models in clinical practice, it is very important to assess their predictive performance before implementation. To this end, as introduced in Chapter 1, the time-dependent ROC curve and the corresponding AUC are the two most popular metrics for assessing the predictive accuracy of a diagnostic score derived from survival models.

During the last decades, for right censored time-to-event data, several authors have proposed estimation methods for time-dependent ROC curve and the corresponding AUC that takes censoring into account. Heagerty et al. (2000) introduced two methods for time-dependent ROC curve. The first one is based on the Kaplan-Meier estimator whereas the second is based on the nearest neighbor estimation (NNE) method. Chambless and Diao (2006) proposed an alternative method to the Kaplan-Meier approach. This is based on a recursive calculation of the ordered times of the event. The advantages of this method is that the sensitivity estimator is monotone, bounded in [0, 1] and does not involve any smoothing parameter. Uno et al. (2007) and Hung and Chiang (2010) independently proposed inverse probability of censoring weighting (IPCW) method to correct the Kaplan-Meier estimate of the sensitivity. Blanche et al. (2013) modified the IPCW and proposed a method called conditional IPCW. The advantage of the latter over IPCW is that it does not assume independence between the prognostic tool and the censoring time. Martínez-Camblor et al. (2016) and Li et al. (2018) independently proposed another method for ROC curve estimation. In their method the missing status indicator for the censored subjects is replaced by weights obtained from the conditional survival function. However, their method was depending on how the conditional survival function is estimated. Martínez-Camblor et al. (2016) used a semiparametric Cox model and Kaplan-Meier estimator for conditional survival function estimation, whereas Li et al. (2018) used a weighted Kaplan-Meier estimator. In this paper we use the estimator proposed by Li et al. and hereafter we call it nonparametric simple (NPS) estimator. The advantage of this method is that it

uses the available information more efficiently than many (if not all) other methods available in the literature like the ICPW and the NNE methods. Moreover this method is relatively insensitive to the choice of the smoothing parameter. More details about this method and the results of a numerical study demonstrating its superiority can be found in Li et al. (2018). Many other methods to estimate the time-dependent AUC exist in the literature, see, for example, Martínez-Camblor and Pardo-Fernández (2018) and the references given therein. For a comprehensive summary of the time-dependent ROC curve estimation methods the reader is referred to Blanche et al. (2013) and Kamarudin et al. (2017).

The estimators of the time-dependent AUC presented above were developed under the assumptions of classical survival models such as non-informative censoring and correctly specified model. Schmid and Potapov (2012) conducted some extensive simulations to study the effects of model misspecification and nonrandom censoring on the behaviors of the estimators of the classification accuracy. They found that violations of these assumptions may generate non-ignorable bias. Schmid et al. (2013) also conducted a similar study to assess the validity of time-dependent AUC estimators and they indicated that violations of these regularity assumptions (i.e., the proportional hazards and the random censoring assumptions) may result in over-optimistic estimates of prediction accuracy.

In the classical survival models, proportional hazards and random censoring are not the only assumptions. Another strong assumption is that all subjects of the study population are susceptible to the event of interest and will eventually experience the event if the follow-up period is sufficiently long. This assumption may not be valid in many situations. In oncology, for example, if the treatment is successful (which might be the case sometimes) the subject will be cured and hence will not have a recurrent event even if the follow-up period is large. This leads to an infinite time to event for cured subjects. In order to obtain valid estimates and inference, it is then necessary to take this into account when modeling survival data. During the last decades, different methods for modeling survival data with cure fraction that yield valid regression coefficient estimates and inferences were proposed. For a recent review on cure models see, for example, Amico and Van Keilegom (2018b) and Taweab and Ibrahim (2014).

Many studies have shown that fitting standard survival models when the data include cured subjects might result in invalid inference (Taweab and Ibrahim; 2014). To the best of our knowledge, no study has yet addressed the effect of the presence of a cure fraction on the properties of the estimates of the time-dependent AUC. Our main aim in this chapter is to address this issue by investigating the following two questions: 1. Is it still valid to use the NPS estimator when there are cured subjects in the study population? The reasons to use this method is that it is a non-parametric method, it efficiently uses the available information and it out-

performs many of other existing methods including the ICPW method (Li et al.; 2018). 2. What are the consequences of deriving a prognostic tool from a survival model that does not acknowledge the presence of cured subject? More precisely, we would like to see what is the cost, in terms of optimality/efficiency, of using a survival working model that do not acknowledge the presence of cured subjects. This question is motivated from the fact that in the current practice, even though a proportion of subjects are not susceptible to the event of interest, it is common to use classical survival models such as the Cox proportional hazards model to derive prognostic tools. So the main question here is about the robustness (to cure) of the classical survival model, which is the most commonly used in practice, as a working model for deriving prognostic tools.

This chapter is organized in five sections including the introduction. Section 2.2 introduces the methodology we used in this study and also define and discuss time-dependent classification accuracy in the presence of cure fraction. In Section 2.3, the simulation design and results are presented and in Section 2.4, we provide real example using data from breast cancer study. Finally, discussions and conclusions are presented in Section 2.5.

2.2 Method

In this section, we first briefly describe the nonparametric simple time-dependent ROC curve and the AUC estimation in the presence of cure fraction. We afterwards give a short description regarding prognostic tools derivation and survival models with cure rate.

Let's assume that a data set of size n is available in order to assess the predictive accuracy of a prognostic tool of interest. More specifically, the data will consist of $\{(Y_i, \Delta_i, \mathbf{X}_i); i = 1, 2..., n\}$ where $Y_i = \min(T_i, C_i)$ is the observed time, T_i the time to event, C_i the censoring time and $\Delta_i = I(T_i \leq C_i)$ the censoring indicator. $\mathbf{X}_i = (X_{1i}, \ldots, X_{di})$, $d \geq 1$, is a d-dimensional vector of patient-specific covariates thought to have a potential influence on T. The event status at time t is denoted by $D_i(t) = I(T_i \leq t)$ with 1 indicating that the event happened by time t. The prognostic tool M, for which the predictive ability is to be investigated, can be any univariate function of $\mathbf{X}$. This function can be either known (e.g., subjective summary rule of expert physicians) or derived (estimated) from a working statistical model fitted to the observed data. More details about prognostic tools derivation are given in Section 2.2.2.

2.2.1 The estimation of ROC curve and AUC in survival models: The nonparametric simple method

The NPS method was proposed for estimating the ROC curve and AUC after reformulating the method of Heagerty et al. (2000) by introducing a weight variable representing the event status of the subject during follow-up (Li et al.; 2018; Martínez-Camblor et al.; 2016). The idea behind this method is similar to the idea of imputation in the context of missing data. They proposed to impute the event status in censored subjects using the survival probability conditional to the observed data. More specifically, the weights are given by:

$$W(t) = \mathrm{P}(T \leq t|Y, \Delta, M).$$

From this, we have two cases: (i) The event is observed (i.e., $\Delta = 1$) $\Leftrightarrow Y = T \leq C$ and (ii) The subject is censored (i.e., $\Delta = 0$) $\Leftrightarrow Y = C < T$.

In the first case,

$$\mathrm{P}(T \leq t|Y, \Delta = 1, M) = I(Y \leq t).$$

In the second case, the weight is given by

$$\begin{aligned} P(T \leq t|M, \Delta = 0, Y) &= P(T \leq t|M, C = Y, T > Y) \\ &= P(T \leq t|M, T > Y) \\ &= 1 - P(T > t|M, T > Y) \\ &= 1 - \frac{P(T > \max(t, Y)|M)}{P(T > Y|M)} \\ &= \begin{cases} 0 & \text{if } Y \geq t \\ 1 - \frac{S(t|M)}{S(Y|M)} & \text{if } Y < t. \end{cases} \end{aligned}$$

This development is true under the usual assumption that T and C are independent given M. By combining the above two cases, the weight can be written as $W(t) = \mathrm{P}(T \leq t|Y, \Delta, M) = \left[1 - (1 - \Delta)\frac{S(t|M)}{S(Y|M)}\right] I(Y \leq t)$, where $S(t|M) = \mathrm{P}(T > t|M)$ denotes the conditional survival distribution of T given the prognostic tool M, which can be estimated using the weighted Kaplan-Meier estimator of Beran with either nearest neighbor or kernel weights; see Beran (1981).

For a given time horizon t and given a sample (Y_i, Δ_i, M_i) of size n, the time-dependent TPR and FPR are estimated as follows:

$$\widehat{TPR}_t(m) = \hat{\mathrm{P}}(M > m|T \leq t) = \frac{\sum_i \hat{W}_i(t) I(M_i > m)}{\sum_i \hat{W}_i(t)},$$

$$\widehat{FPR}_t(m) = \hat{\mathrm{P}}(M > m|T > t) = \frac{\sum_i (1 - \hat{W}_i(t)) I(M_i > m)}{\sum_i (1 - \hat{W}_i(t))},$$

where $I(.)$ is the indicator function and $\hat{W}_i(t) = \left[1 - (1 - \Delta_i)\frac{\hat{S}(t|M_i)}{\hat{S}(Y_i|M_i)}\right] I(Y_i \leq t)$ is the estimate of the weight function defined above, where $\hat{S}(.|M_i)$ is the Beran's estimate of $S(.|M_i)$ given in equation (1.2).

The time-dependent ROC curve can be obtained by plotting the sensitivity versus 1-specificity for all possible cutoff values m of the prognostic tool M and the corresponding time-dependent AUC can be calculated by using numerical integration such as the trapezoidal method. An alternative way for calculating the AUC without involving numerical integration is to use the Mann-Whitney U statistic as it has close connection with the AUC. In fact, Hanley and McNeil (1982) showed that the AUC is equivalent to the probability that a randomly selected subject with the event of interest has a higher prognostic tool than a randomly chosen subject without the event of interest. Using this the AUC estimate can be obtained using the following formula:

$$\widehat{AUC}_t = \frac{\sum_i \sum_j \hat{W}_i(t)(1 - \hat{W}_j(t))[I(M_i > M_j) + 0.5 \times I(M_i = M_j)]}{\sum_i \sum_j \hat{W}_i(t)(1 - \hat{W}_j(t))}, \qquad (2.1)$$

where i and j are the indexes of two independent subjects (Li et al.; 2018).

In the presence of cure fraction the aforementioned time-dependent classification accuracy measures can be defined as follows: Let $B_i = I(T_i < \infty)$ be the latent variable indicating the uncure status taking values 1 or 0 depending on whether the subject i is uncured or not. In this work, for a given time horizon t, the case and the control groups are taken to be $\{B_i = 1 \text{ and } T_i \leq t\} = \{T_i \leq t\}$ and $\{B_i = 0 \text{ or } T_i > t\} = \{T_i > t\}$, respectively. With these definitions one gets the so-called *cumulative* sensitivity and *dynamic* specificity, see equation (1.12), that are largely used in the classical survival analysis (without a cure fraction). The resulting ROC, see equation (1.13), is known as the *cumulative/dynamic* (C/D) ROC; see Heagerty et al. (2000) and Heagerty and Zheng (2005) for more detail. This suggests that the NPS estimators described above can be used to estimate the time-dependent classification accuracy measures in the presence of cure fraction.

Note that since the Beran estimator is the main ingredient of NPS estimator, the consistency of the latter holds if the former is consistent. Without a cure fraction, the consistency and asymptotic normality of Beran's estimator were established by several researchers, just to mention a few Beran (1981), Dabrowska (1989), González-Manteiga and Cadarso-Suárez (1994) and Van Keilegom and Veraverbeke (1997). With a cure fraction, due to the identifiability issues related to censoring and cure, for consistency to hold, we need, among other technical assumptions, that the support of the censoring variable to be not contained in the support of T. More precisely, we need that $\tau_u < \tau_G(M)$, with probability one, where $\tau_u = \sup_m\{\inf\{t : S_u(t|m) = 0\}\}$ and $S_u(t|m) = \mathrm{P}(T \geq t|M = m, T < \infty)$

is the conditional survival function of the uncured subjects, and $\tau_G(m) = \inf\{t : G(t|m) = 0\}$ where $G(t|m)$ is conditional survival function of the censoring time. For this reason, one should restrict the analysis to a time horizon t not larger than τ_u; for more details see Xu and Peng (2014) and López-Cheda et al. (2017).

2.2.2 Derivation of prognostic tools in survival analysis

As stated in the previous section, a prognostic tool can be either known or derived. An example of the first case is the Glasgow Coma Scale (GCS) routinely used in emergency medicine; see, e.g., Baršić et al. (1996). The GCS measures the severity of coma in patients admitted to intensive care unit and is usually used to predict the 24 or 48 hours in-hospital risk of death. It is a straightforward test that assesses the patients ability to open their eyes, move and speak. The GCS test assesses three main vital components: The motor responses ranging from 1 to 6, the verbal responses ranging from 1 to 5 and the eye opening ranging from 1 to 4 resulting in a total score ranging from 3 to 15. In the second case, no known prognostic tool is available. The common practice is then to derive a prognostic tool by combining information from different potential predictors. Several researchers proposed optimal ways to combine information from different sources (see, for example, McIntosh and Pepe (2002); Qin and Zhang (2010); Cai and Cheng (2008)). The most common way is by fitting a working survival model $\mathcal{W}$ to a set of data collected on the event of interest and potential predictors. An example of this is the United Kingdom prospective diabetes study (UKPDS) risk engine used to predict the 10 years risk of developing coronary heart disease in patients with type 2 diabetes; see Stevens et al. (2001). This prognostic tool was derived by fitting a parametric survival model ($\mathcal{W}$) to data prospectively collected on type 2 diabetes patients.

It has been shown that, given a set of predictors $\mathbf{X} = (X_1, \ldots, X_d)$, the conditional probability $P(T \leq t|\mathbf{X})$, or any monotonic increasing transformation of it, is the optimal prognostic tool for predicting $I(T \leq t)$. Optimal here means that it maximizes the AUC (McIntosh and Pepe; 2002). For example, if the data are generated from the model

$$h(T) = \alpha + \boldsymbol{\beta}'\boldsymbol{X} + \epsilon, \tag{2.2}$$

where h is an increasing function and ϵ is a random error independent of $\boldsymbol{X}$, then the prognostic tool $M = \boldsymbol{\beta}'\boldsymbol{X}$ is optimal. This is also the case for the classical Cox model

$$S^{\mathcal{COX}}(t|\boldsymbol{X}) = \exp(-\Lambda_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{X})), \tag{2.3}$$

where $\Lambda_0(t)$ is the cumulative baseline hazard function.

In practice, the parameters of the working model used to define the prognostic tool, e.g., $\boldsymbol{\beta}$ in the Cox model, are unknown and need to be estimated form a training/learning data set, using, for example, (partial)likelihood method. In the

following, we denote by $\hat{M}^{\mathcal{W}}$ the estimated prognostic tool based on the working model $\mathcal{W}$, e.g., $\hat{M}^{\mathcal{COX}} = \mathbf{X}^T\hat{\boldsymbol{\beta}}^{\mathcal{COX}}$, where $\hat{\boldsymbol{\beta}}^{\mathcal{COX}}$ is the partial likelihood estimator of $\boldsymbol{\beta}$ in the Cox model. The prognostic tool $\hat{M}^{\mathcal{COX}}$ is optimal when the Cox model holds but may have poor accuracy when the proportional hazards assumption fails, for more details see Cai and Cheng (2008) and Zheng et al. (2006).

2.2.3 Survival models with cure

As presented in the previous chapter, to analyze time-to-event data with cure fraction, survival models, also known as cure models, have been widely developed to take the presence of cured fraction into account. One of the most popular classes of these models is the mixture cure model (MCM), which introduced by Boag (1949). This model assumes that the population is a mixture of cured individuals that will never experience the event of interest, and non-cured individuals, who may experience it. The mixture cure model is a combination of two models, one for the probability of cure sometimes called "the incidence model" and one for the event time of uncured subjects sometimes called "the latency model". The survival function in the mixture cure model is given as

$$S^{\mathcal{MCM}}(t|\mathbf{X}) = p(\mathbf{X})S_u(t|\mathbf{X}) + (1 - p(\mathbf{X})), \tag{2.4}$$

where $\mathbf{X}$ is the vector of covariates, $S^{\mathcal{MCM}}(t|\mathbf{X}) = \mathrm{P}(T > t|\mathbf{X})$ is the conditional population survival function and $p(\mathbf{X}) = \mathrm{P}(B = 1|\mathbf{X})$ is the conditional probability of being uncured. The function $S_u(t|\mathbf{X}) = \mathrm{P}(T > t|\mathbf{X}, B = 1)$ is the conditional survival function of the susceptible subjects. This function can be assumed to belong to a parametric or a semiparametric distribution family like, for example, the widely used family of Cox proportional hazards models.

Another important class of cure model called the promotion time model, also known as "proportional hazard cure model" or "bounded cumulative hazard model" (BHM), was later introduced by Yakovlev et al. (1993). This model was strongly motivated by a biological mechanism assuming the number of cancer cells remaining after treatment to follow a Poisson distribution (Yakovlev et al.; 1993; Taweab and Ibrahim; 2014). The survival function in the BHM is given as

$$S^{\mathcal{BHM}}(t|\mathbf{X}) = \exp\{-\theta(\mathbf{X})F(t)\}, \tag{2.5}$$

where $\mathbf{X}$ is the vector of covariates, $F(t)$ is a proper cumulative distribution function and $\exp\{-\theta(\mathbf{X})\}$ is the probability of being cured.

For the reasons explained above and for easiness of interpretation, the prognostic tools considered in all the following sections will be in the form $1 - S^{\mathcal{W}}(t|\boldsymbol{X})$, where $\mathcal{W}$ is one of the working models under study, i.e. COX, MCM or BHM, and $S^{\mathcal{W}}$ is the corresponding estimated survival function.

2.3 Simulations

The main objective of this simulation study is to assess the validity of the NPS estimator when there are cured subjects in the population under study. We also investigated the loss in terms of predictive ability when deriving prognostic tools from survival models that do not acknowledge the presence of cured subjects. Details of these simulations are provided in the upcoming subsections.

To generate the improper survival time T we use the following equation

$$T = T^u + \frac{1-B}{B},$$

where T^u is a proper survival time (i.e., the time to event variable of the uncured subpopulation) and B is the cure indicator generated from a Bernoulli distribution with probability p. The distribution of T^u and the values of p will change according to our simulation settings and will be specified in the following subsections. In all our simulations, we used the NPS method implemented in `R` package `tdROC` (Li and Wu; 2016) with the default value (i.e. 0.1) for the smoothing parameter that is needed to calculate the weighted Kaplan-Meier estimate with 0/1 nearest neighbor weights which is given by $I\{-\lambda < \hat{F}(M_i) - \hat{F}(M_j) < \lambda\}$, where $2\lambda \, \epsilon \, (0,1)$ represents the percentage of observations that is included in each neighborhood and $\hat{F}(.)$ is an estimate of the distribution function of M (Heagerty et al.; 2000). We also tried different values of λ and observed minimal differences in the performance measures of the AUC estimates. This confirms that the NPS method is relatively insensitive to the choice of the smoothing parameter as shown in Li et al. (2018).

2.3.1 The validity of the NPS estimator under cure: The case of a known prognostic tool

The objective of this section is to investigate the performance of the NPS estimator of the time dependent AUC when the risk tool is known or fixed. The performance of the estimator will be assessed by calculating the percentage bias and the mean square error. Two scenarios were considered. In the first scenario, we assume that the probability of cure does not depend on the covariates whereas the second scenario assumes this probability to be dependent on the covariates. For both of these scenarios the prognostic tool is defined as $M = 1 - S^{\mathcal{MCM}}(t|X)$ and the data were generated from a mixture cure model as follows :

Scenario 1: As in Li et al. (2018), Heagerty and Zheng (2005) and Martínez-Camblor et al. (2016), we generate $(X, \log(T^u), \log(C))$ from a trivariate normal

distribution (MVN). More precisely

$$\begin{bmatrix} X_i \\ \log\,(T_i^u) \\ \log\,(C_i) \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \\ \mu_c \end{bmatrix}, \quad \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right),$$

where -0.9, so that a higher value of the covariate X is an indication of a shorter duration T. To control the censoring and cure proportions, different values of μ_c and p (uncured probability) were considered. These values together with the chosen time horizon t and the percentages of cure (%cure) and $AUC_M(t)$ can be found in Table 2.1 (Scenario 1). Note that in this scenario the uncured probability p does not depend on the covariate X.

Scenario 2: The data are generated from a logistic-exponential MCM, where the probability of being uncured is given by:

$$p(X) = \frac{\exp(\gamma_0 + X)}{1 + \exp(\gamma_0 + X)},$$

and the conditional survival function of the uncured subjects is given by:

$$S_u(t|X) = \exp(-\exp(-3X)t).$$

The covariate X is assumed to follow a standard normal distribution. The censoring time was generated from an exponential distribution with parameter λ. We consider different values for the parameters λ and γ_0 to get different censoring and cure fractions. These parameters are summarized in Table 2.1 (Scenario 2). Note that contrary to scenario 1, the uncured probability p depend here on X. The true $AUC_M(t)$ in Table 2.1 was calculated empirically, from a sample $(T_i, M_i), i = 1, \ldots, 4 \times 10^6$, generated as shown in scenarios 1 and 2, by using equation (2.1) with $\hat{W}_i = I(T_i \leq t)$.

For each scenario, we generated $N = 2000$ samples of size $n = 200$ and 400 and evaluated the performance of the AUC estimator using the empirical percent bias (% bias) and empirical mean squared error (MSE), given, respectively, by

$$\%\text{ bias} = \frac{\hat{\text{E}}(\hat{AUC}_M(t)) - AUC_M(t)}{AUC_M(t)} \times 100, \quad \text{MSE} = N^{-1} \sum_{j=1}^{N} (\hat{AUC}_{M,j}(t) - AUC_M(t))^2,$$

where $\hat{\text{E}}(\hat{AUC}_M(t)) = N^{-1} \sum_{j=1}^{N} \hat{AUC}_{M,j}(t)$ and $\hat{AUC}_{M,j}(t)$ is the NPS estimate calculated from the j-th ($j = 1, \ldots, N$) simulated data $(Y_{i,j}, \Delta_{i,j}, M_{i,j})$, $i = 1, \ldots, n$,

Table 2.1: Parameters used to generate the data for Scenario 1 and 2 and the resulting percentages of cure (%cure) and true values of $AUC_M(t)$.

		Scenario 1		Scenario 2		
t	%cure	μ_c	$AUC_M(t)$	λ	γ_0	$AUC_M(t)$
0.2	20	0.94	0.966	0.0550	1.65	0.902
0.8	20	0.00	0.886	0.4400	1.65	0.857
2.5	20	0.00	0.736	0.4400	1.65	0.823
0.2	30	1.45	0.964	0.0100	1.00	0.893
0.8	30	0.25	0.863	0.1740	1.00	0.838
2.5	30	0.25	0.691	0.1740	1.00	0.795
0.2	40	3.80	0.961	0.0001	0.50	0.896
0.8	40	0.60	0.844	0.0620	0.50	0.826
2.5	40	0.60	0.660	0.0620	0.50	0.780

Table 2.2 shows the results obtained for the sample sizes $n = 200$ and $n = 400$, censoring rates 40% and 60%, cure rates 20%, 30% and 40% and time horizons $t = 0.2$ and $t = 0.8$. From these results, it can be seen that, for both scenarios, the NPS AUC estimator performed very well, with small percent bias (less than 0.3%) and small MSE (less than 0.9×10^{-3}). Furthermore, as the sample size increased from 200 to 400, the MSE was reduced by around 50%, which indicates that the NPS estimator is consistent for data with cured subjects. In addition, the bias and MSE decreases when the strength of association between the event time and the marker increase. The results also show that, as the percentage of censored and cure rate increases, both the percent bias and the MSE increase. A similar effect is also observed when we increase the time horizon t. Moreover, the effect of the censoring percentage increases as the time horizon t increase. The possible reason for the MSE difference increase with time horizon t between 40 and 60 percent censoring can be explained by the proportion of subjects with the observed time larger than t (i.e., $P(Y > t)$) difference increase between 40 and 60 percent censoring when the time horizon t increases. From these results, we can conclude that, when the prognostic tool is known, whether the probability of cure depends on covariate X or not, the simple method gives very good results. This may be explained by the fact that, as we mention before, the NPS relies on Beran's estimator which is consistent even in the presence of a cure fraction in the data (López-Cheda et al.; 2017). However, care should be taken for the choice of the time horizon t as Beran's estimator is not consistent everywhere.

Table 2.2: % bias and MSE of the NPS AUC estimate for data generated according to Scenario 1 and 2.

				Scenario 1		Scenario 2	
t	%cure	%cens	n	% bias	MSE ($\times 10^{-3}$)	% bias	MSE ($\times 10^{-3}$)
0.2	20	40	200	-0.042	0.358	0.076	0.419
0.2	20	40	400	0.014	0.167	0.098	0.200
0.2	20	60	200	-0.064	0.368	0.110	0.435
0.2	20	60	400	0.005	0.180	0.105	0.207
0.2	30	40	200	-0.108	0.444	-0.043	0.576
0.2	30	40	400	-0.054	0.207	0.018	0.287
0.2	30	60	200	-0.111	0.450	-0.046	0.578
0.2	30	60	400	-0.059	0.214	0.015	0.291
0.2	40	40	200	0.001	0.615	-0.108	0.609
0.2	40	40	400	0.032	0.246	-0.052	0.293
0.2	40	60	200	-0.009	0.617	-0.112	0.610
0.2	40	60	400	0.024	0.250	-0.053	0.294
0.8	20	40	200	0.042	0.534	0.112	0.565
0.8	20	40	400	0.092	0.266	0.088	0.285
0.8	20	60	200	0.206	0.703	0.186	0.669
0.8	20	60	400	0.194	0.345	0.139	0.340
0.8	30	40	200	0.062	0.652	-0.002	0.636
0.8	30	40	400	0.133	0.318	0.079	0.336
0.8	30	60	200	0.210	0.784	0.029	0.684
0.8	30	60	400	0.233	0.378	0.098	0.357
0.8	40	40	200	-0.022	0.757	-0.171	0.847
0.8	40	40	400	0.024	0.357	-0.139	0.432
0.8	40	60	200	0.004	0.853	-0.164	0.850
0.8	40	60	400	0.096	0.385	-0.140	0.437
2.5	20	40	200	-0.347	1.834	-0.095	0.747
2.5	20	40	400	-0.290	0.916	-0.079	0.397
2.5	20	60	200	-0.772	3.705	0.057	1.241
2.5	20	60	400	-0.789	1.781	0.095	0.665
2.5	30	40	200	-0.191	1.611	0.029	0.760
2.5	30	40	400	-0.158	0.791	0.049	0.398
2.5	30	60	200	-0.818	3.009	0.036	0.938
2.5	30	60	400	-0.792	1.403	0.044	0.483
2.5	40	40	200	0.023	1.431	0.005	0.878
2.5	40	40	400	0.123	0.693	0.050	0.448
2.5	40	60	200	-0.480	2.201	-0.023	0.939
2.5	40	60	400	-0.537	1.080	0.031	0.480

2.3.2 The validity of the NPS estimator under cure: The case of an estimated prognostic tool

The objective of this section is to assess the performance of the NPS estimator in the case where the prognostic (risk) tool needs to be estimated from the data using a given working survival model. In order to avoid optimistic bias of future performance, we followed the established rule which consists in estimating (or deriving) the risk tool from a learning set and then evaluating this tool and comparing its prediction accuracy on an independent validation set. This method helps to avoid overoptimistic prognoses and it is more appropriate than using the same data for both learning and validation (Schmid and Potapov; 2012; Krzanowski and Hand; 2009).

Two approaches were used to answer this objective. In the first approach we compared $\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)$ to $AUC_{M^{\mathcal{W}}}(t)$ where $\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)$ is a random variable representing the NPS AUC estimate of the risk tool $\hat{M}^{\mathcal{W}}$ derived from fitting a working survival model $\mathcal{W}$ to a finite survival data set and $AUC_{M^{\mathcal{W}}}(t)$ is the true AUC of the risk tool $M^{\mathcal{W}}$ where $M^{\mathcal{W}}$is derived from fitting a working survival model $\mathcal{W}$ to a huge survival data set (population). The percent bias and MSE according to approach 1 are defined as follows:

$$\%\text{bias} = \frac{\mathrm{E}(\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)) - AUC_{M^{\mathcal{W}}}(t)}{AUC_{M^{\mathcal{W}}}(t)} \times 100, \quad \text{MSE} = \mathrm{E}[\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t) - AUC_{M^{\mathcal{W}}}(t)]^2. \tag{2.6}$$

In the second approach we compared $\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)$ to $AUC_{\hat{M}^{\mathcal{W}}}(t)$ where $\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)$ is defined as above and $AUC_{\hat{M}^{\mathcal{W}}}(t)$ is the true population AUC of the risk tool $\hat{M}^{\mathcal{W}}$ where $\hat{M}^{\mathcal{W}}$ is derived by fitting a working survival model $\mathcal{W}$ to a finite survival data set. The percent bias and MSE according to approach 2 are defined as follows:

$$\%\text{bias} = \frac{\mathrm{E}(\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)) - AUC_{\hat{M}^{\mathcal{W}}}(t)}{AUC_{\hat{M}^{\mathcal{W}}}(t)} \times 100, \quad \text{MSE} = \mathrm{E}[\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t) - AUC_{\hat{M}^{\mathcal{W}}}(t)]^2. \tag{2.7}$$

In approach 2, we try to eliminate the error due to the estimation of the marker. By doing so, we focus only on the performance of the AUC estimator (given an estimated marker).

Approach 1: Comparing $\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)$ to $AUC_{M^{\mathcal{W}}}(t)$.

For the first approach, we proceeded with the following simulation plan:

(i) We generated a learning sample of size n_l *i.e.* $(Y_i, \Delta_i, \mathbf{X}_i)$, $i = 1, \ldots, l$ using a MCM.

(ii) From the generated data in (i), we fitted a working survival model $\mathcal{W}$, extracted the required estimated parameters and derived an estimated risk tool

$\hat{M}^{\mathcal{W}}(\mathbf{X}) \equiv \hat{M}^{\mathcal{W}} = 1 - \hat{S}^{\mathcal{W}}(t|\mathbf{X})$, where $\hat{S}^{\mathcal{W}}$ is the estimated conditional survival function of T based on $\mathcal{W}$.

(iii) Using the same MCM as in step (i), we generated a validation sample, $(Y_i, \Delta_i, \mathbf{X}_i)$, of size n_v and calculated the risk tool $\hat{M}_i^{\mathcal{W}} = \hat{M}^{\mathcal{W}}(\mathbf{X}_i)$, $i = 1, \ldots, n_v$.

(iv) Based on $(Y_i, \Delta_i, \hat{M}_i^{\mathcal{W}})$, $i = 1, \ldots, n_v$, we calculated $\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)$, the NPS estimator of the estimated risk tool $\hat{M}^{\mathcal{W}}$.

In step (ii) of the simulations plan above, two working survival models $\mathcal{W}$ were fitted: The Cox proportional hazards model (COX) and the mixture cure model (MCM). The MCM model used to generate the data in steps (i) and (iii) is a mixture of logistic and exponential where

$$p(\mathbf{X}) = \frac{\exp(\gamma_0 + \gamma_1 X_1 + \gamma_2 X_2)}{1 + \exp(\gamma_0 + \gamma_1 X_1 + \gamma_2 X_2)},$$

is the probability of uncured and

$$S_u(t|\mathbf{X}) = \exp(-\exp(\beta_1 X_1 + \beta_2 X_2)t),$$

is the conditional survival function of the uncured subjects and the covariates $X_1 \sim N(0,1)$ and $X_2 \sim Unif[0,1]$. The censoring time was generated from an exponential distribution with parameter λ. Note that the covariates (X_1, X_2) affect both the probability of being uncured and the survival function of uncured subjects. We considered two cases in parameter settings:

$$\begin{array}{l} \text{Case 1: } (\beta_1, \beta_2) = (\gamma_1, \gamma_2) = (-1, -1), \\ \text{Case 2: } (\beta_1, \beta_2) = -(\gamma_1, \gamma_2) = (1, 1). \end{array}$$

Note that in step (ii), we could have used the linear predictor function $\mathbf{X}'\boldsymbol{\beta}$ as a potential prognostic tool instead of the distribution function $M^{\mathcal{W}} = 1 - S^{\mathcal{W}}(t|\mathbf{X})$. This will indeed lead to the same AUC when $\mathcal{W} = COX$ or when $\mathcal{W} = MCM$ but only under case 1. This is, however, not true for case 2 since the survival function is not monotonic in $\mathbf{X}'\boldsymbol{\beta}$ in that case.

To control the censoring and cure proportions, we considered different values for the parameters λ and γ_0. These parameters are summarized in Table 2.3.

Table 2.3: Parameters used to generate the data for case 1 and case 2 and the resulting percentages of cure (%cure) and true values of $AUC_{M^{\mathcal{MCM}}}(t)$.

		Case 1			Case 2		
t	%cure	λ	γ_0	$AUC_{M^{\mathcal{MCM}}}(t)$	λ	γ_0	$AUC_{M^{\mathcal{MCM}}}(t)$
1	20	0.200	2.17	0.826	0.130	1.170	0.701
2	20	0.730	2.17	0.836	0.500	1.170	0.691
1	30	0.100	1.53	0.830	0.050	0.520	0.662
2	30	0.560	1.53	0.832	0.310	0.520	0.654
1	40	0.001	1.00	0.833	0.001	0.001	0.633
2	40	0.380	1.00	0.826	0.180	0.001	0.629

For each case and each combination of the parameters in Table 2.3, we repeated steps (i)-(iv) described above $N = 2000$ times with sample sizes $n = n_l = n_v = 200, 400, 600$ and evaluated the bias and MSE. To calculate the $AUC_{Z^{\mathcal{MCM}}}(t)$, the true optimal AUC, we generated a sample $(T_i, \mathbf{X}_i)$ of size 4×10^6 from the MCM and then calculated the AUC of $M = 1 - S^{\mathcal{MCM}}(t|\mathbf{X})$ empirically, from $(T_i, M_i^{\mathcal{MCM}})$, using equation (2.1) with $\hat{W}_i = I(T_i \leq t)$. For the $AUC_{M^{cox}}(t)$, the *pseudo* true AUC, we run $1-3$ once with $n_l = n_v = 4 \times 10^6$ and then calculated $AUC_{M^{cox}}(t)$ empirically from $(T_i, M_i^{\mathcal{COX}})$. Note that the MSE was calculated according to the formula MSE $= N^{-1}\sum_{j=1}^{N}(\hat{AUC}_{\hat{M}_j^{\mathcal{W}},j}(t) - AUC_{M^{\mathcal{W}}}(t))^2$, where $\hat{AUC}_{\hat{M}_j^{\mathcal{W}},j}(t)$ is the NPS estimate calculated from the j-th ($j = 1, \ldots, N$) validation sample using the estimated prognostic tool $\hat{M}_j^{\mathcal{W}}$ calculated from the j-th learning sample.

We also repeated all computations in steps (i)-(iv) using a BHM (Bounded Hazard Model) instead of the MCM. More precisely, we generated the survival times from $S(t|\mathbf{X}) = \exp(-\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)F(t))$, where $(\beta_1, \beta_2) = (1, 0.6)$, $F(t) = 1 - \exp(-0.2t)$, $X_1 \sim N(0, 1)$ and $X_2 \sim Unif[0, 1]$. The censoring time was generated from an exponential distribution with parameter λ. We varied the values of λ and β_0 to get the same censoring and cure rates as the one obtained for the MCM. We also used the same time horizon ($t = 1, 2$) as for the MCM. For the sake of brevity, the results of the BHM are not reported here because they were very similar to the results obtained with the MCM in case 1.

In Figures 2.1 and 2.2 we provide the boxplots of the resulting estimates for case 1 and case 2, respectively, whereas in Table 2.4 we show the percent bias and the empirical MSE of these estimates for both case 1 and case 2.

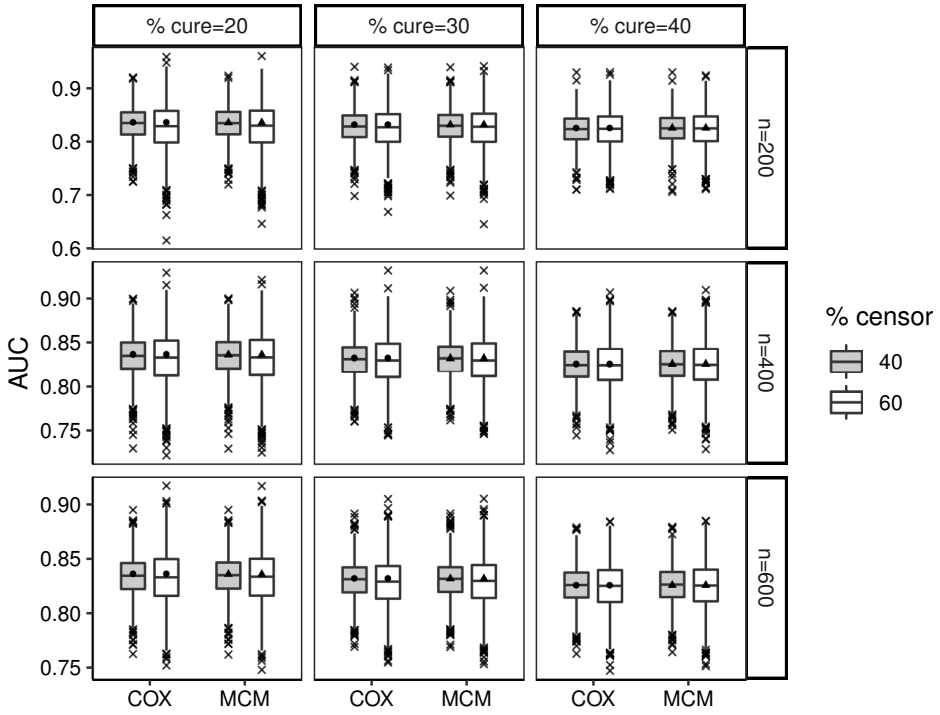

Figure 2.1: Boxplots of $\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)$ in Case 1 for the time horizon $t = 2$. The solid circles and triangles indicates the "true" AUC values $AUC_{M^{\mathcal{W}}}(t)$ of the Cox and mixture cure working models, respectively.

Let us first look at case 1. The boxplots in Figure 2.1 and the results in Table 2.4 reveal that the prognostic tool estimate based on the Cox model and on the mixture cure model result in almost exactly the same estimates and so the same performances. This is true for all cure rates, censoring rates and sample sizes. We can also see that the bias and MSE increase when the censoring and/or the cure rates increase. In contrast, the bias and MSE decrease as the sample size increase. We obtained exactly the same conclusions when we analyzed the results of the BHM (not shown here). We can therefore conclude that under case 1, the NPS method still provides valid and consistent estimates of the true AUC even in the situation where the prognostic tool is derived from a misspecified model such as the Cox proportional hazard model. This may be explained by the fact that, in this case, the misspecified Cox model estimates the vector of parameters $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ consistently. This is indeed the case for BHM as it has been shown in Portier et al. (2017).

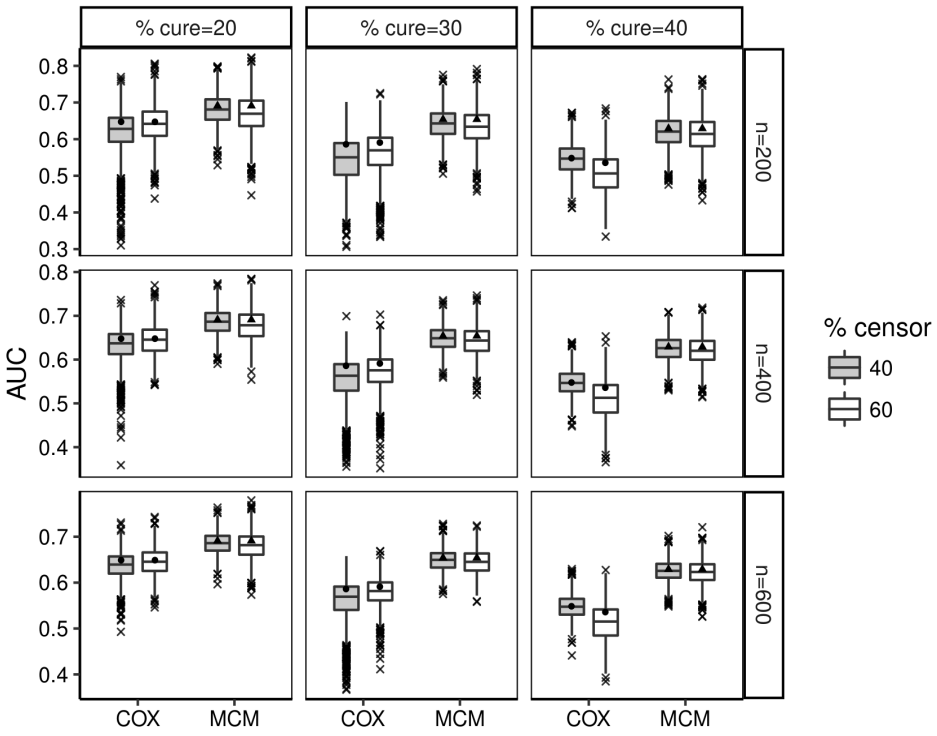

Figure 2.2: Boxplots of $\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)$ in Case 2 for the time horizon $t = 2$. The solid circles and triangles indicates the "true" AUC values $AUC_{M^{\mathcal{W}}}(t)$ of the Cox and mixture cure working models, respectively.

For case 2, the most common situation (the coefficients in the cure probability are different from those in the survival component), the story is completely different. In fact both the results presented in Figure 2.2 and Table 2.4 indicate that deriving (estimating) a prognostic tool from a survival data exhibiting cured subjects using a Cox proportional hazards model results in larger bias and larger MSE in the AUC estimates compared to a mixture cure model for almost all studied cases. This is true for all cure rates, censoring rates, and sample sizes. This bias is noticeable even when the sample size is large ($n = 600$), especially when the cure rate is high. For example, for $t = 1$, when %cure = 40 and %cens = 60, the %bias using COX model is almost ten times larger than the one obtained using MCM. In contrast, the prognostic tool estimate using the mixture cure model generally result in an AUC estimate with very small bias (less than 1.5%) and MSE (less than 0.001) that decrease with an increasing sample size. The possible reason for

the large bias and MSE observed in the AUC estimates when using the misspecified Cox working model, in this case, can be explained by the bias in estimating the survival function for data with a cure fraction. This was well documented in Amico and Van Keilegom (2018b). In their paper, when comparing the MCM and Cox model in a simulation study, the authors concluded that "not taking into account the presence of a cure fraction in survival data has important consequences that may lead to wrong conclusions". In our simulations study, we also investigated the case when the data is generated for the classical Cox model (without cure) and the AUC is estimated using either the Cox or the mixture cure working model. In all the investigated cases, both these models lead to almost exactly the same AUC estimates. For this reason and for sake of brevity, we do not show these results here.

Table 2.4: The percent bias and MSE for Case 1 and Case 2 (Approach 1) MCM is mixture cure model and COX is classical Cox model.

$t = 1$			Case 1				Case 2			
			$n = 200$		$n = 600$		$n = 200$		$n = 600$	
%cure	%cens	Model	% bias	MSE ($\times 10^{-3}$)	% bias	MSE ($\times 10^{-3}$)	% bias	MSE ($\times 10^{-3}$)	% bias	MSE ($\times 10^{-3}$)
20	40	MCM	-0.126	1.000	-0.119	0.320	-0.855	1.561	-0.359	0.517
20	40	COX	-0.175	1.000	-0.154	0.322	-4.894	5.685	-1.816	1.065
20	60	MCM	-0.360	1.383	-0.217	0.415	-1.418	1.960	-0.709	0.664
20	60	COX	-0.336	1.364	-0.242	0.416	-1.209	2.122	-0.251	0.652
30	40	MCM	-0.096	0.965	-0.018	0.300	-1.340	1.802	-0.444	0.574
30	40	COX	-0.206	0.983	-0.059	0.300	-10.65	12.18	-6.000	5.312
30	60	MCM	-0.315	1.248	-0.115	0.384	-2.077	2.148	-0.799	0.706
30	60	COX	-0.339	1.258	-0.159	0.384	-5.995	6.326	-2.294	1.465
40	40	MCM	-0.016	0.861	0.021	0.288	-1.422	2.117	-0.530	0.661
40	40	COX	-0.188	0.885	-0.035	0.290	-0.782	1.993	-0.166	0.664
40	60	MCM	-0.158	1.076	-0.044	0.354	-2.071	2.388	-0.892	0.764
40	60	COX	-0.226	1.086	-0.096	0.357	-11.09	10.54	-8.569	6.946
$t = 2$										
20	40	MCM	-0.180	0.980	-0.156	0.322	-1.486	1.805	-0.715	0.561
20	40	COX	-0.237	0.979	-0.202	0.324	-4.122	4.205	-1.550	0.956
20	60	MCM	-1.008	2.015	-0.353	0.616	-3.154	3.226	-1.514	0.984
20	60	COX	-1.043	2.027	-0.390	0.614	-1.026	2.590	-0.280	0.862
30	40	MCM	-0.292	0.915	-0.125	0.290	-1.798	1.788	-0.764	0.545
30	40	COX	-0.400	0.931	-0.174	0.291	-7.455	6.075	-4.240	2.628
30	60	MCM	-0.678	1.486	-0.297	0.494	-3.162	2.675	-1.381	0.804
30	60	COX	-0.742	1.497	-0.361	0.495	-4.455	4.054	-1.832	1.080
40	40	MCM	-0.117	0.794	0.054	0.283	-1.471	1.848	-0.517	0.554
40	40	COX	-0.289	0.814	-0.005	0.286	-0.412	1.715	-0.086	0.597
40	60	MCM	-0.271	1.172	-0.062	0.424	-2.537	2.525	-0.982	0.702
40	60	COX	-0.353	1.182	-0.128	0.427	-5.450	3.881	-4.292	2.131

Approach 2: Comparing $\hat{AUC}_{\hat{M}^{\mathcal{W}}}(t)$ to $AUC_{\hat{M}^{\mathcal{W}}}(t)$.

For the second approach and to reduce the burden of calculations, we investigated only case 2. We proceeded with the following simulation plan:

(i) We generated a learning sample of size n_l from the MCM-case 2 given above.

(ii) From the generated data, we estimated the working model $\mathcal{W}$ and calculated the prognostic tool $\hat{M}^{\mathcal{W}}(\mathbf{X}) \equiv \hat{M}^{\mathcal{W}} = 1 - \hat{S}^{\mathcal{W}}(t|\mathbf{X})$.

(iii) We generated $B = 1000$ independent validation samples of size n_v from the same MCM-case 2.

(iv) From each validation sample, we calculated the NPS estimator $\hat{AUC}_{j,\hat{M}^{\mathcal{W}}}$, $j = 1, \ldots, B$.

Similar to approach 1, two working survival models were fitted in step 2 namely, the Cox proportional hazard model and the mixture cure model. The specification of the MCM is similar to the one described in approach 1. For each combination of the parameters in case 2 (see Table 2.3), we repeated steps (i)−(iv) described above $N = 200$ times ($k = 1, \ldots, N$) and evaluated, for each k, the percent bias and MSE according to equation (2.7). For a given $\hat{M}_k^{\mathcal{W}}$, $AUC_{\hat{M}_k^{\mathcal{W}}}$, the "true" or population version of the AUC, was calculated from a validation sample of size 10^6. Note that, unlike the simulation design carried out in approach 1 where the prognostic tool and its AUC were estimated for each pair of the simulated learning and validation data sets, here in approach 2, once estimated from a given learning data, the prognostic tool M was kept fixed and its AUC evaluated using B validation data sets. To account for the sampling variability, we repeated this process N times and report the averaged performance measures.

Table 2.5 reports the empirical averaged % bias and MSE, given, respectively, by bias $= \frac{1}{N}\sum_{k=1}^{N} \%\,\text{bias}_k$ and MSE $= \frac{1}{N}\sum_{k=1}^{N} \text{MSE}_k$, where $\text{bias}_k = 100 \times (\frac{1}{B}\sum_{j=1}^{B} \hat{AUC}_{j,\hat{M}_k^{\mathcal{W}}}(t) - AUC_{\hat{M}_k^{\mathcal{W}}}(t))/AUC_{\hat{M}_k^{\mathcal{W}}}(t)$ and $\text{MSE}_k = \frac{1}{B}\sum_{j=1}^{B}(\hat{AUC}_{j,\hat{M}_k^{\mathcal{W}}}(t) - AUC_{\hat{M}_k^{\mathcal{W}}}(t))^2$, for sample sizes $n = n_l = n_v = 200, 400$. We also calculated the resulting averaged AUC estimate given by $\hat{AAUC}_{\hat{M}_k^{\mathcal{W}}} = B^{-1}\sum_{j=1}^{B} \hat{AUC}_{j,\hat{M}_k^{\mathcal{W}}}$, for $k = 1, \ldots, N$, and the averaged true AUC given by $AAUC_{\hat{M}^{\mathcal{W}}} = \frac{1}{N}\sum_{k=1}^{N} AUC_{\hat{M}_k^{\mathcal{W}}}$ and reported them using boxplots in Figure 2.3. As done previously, all the results are stratified according to the percentage of censoring and percentage of cure.

The boxplots in Figure 2.3 and the values in Table 2.5 both show that, when the estimated prognostic tool is held fixed, the NPS AUC estimator provides small bias and MSE, especially for a large sample size ($n = 400$), for both the Cox and the MCM models. This is true for all cure rates, censoring rates and sample sizes. Furthermore, the bias and MSE decrease when the sample size increase for both

the MCM and the Cox working models. We can also see from both results that the estimated prognostic tool based on the Cox working survival model resulted in an AUC estimate that converge to its limiting value $AUC_{\hat{M}^{Cox}}(t)$ despite the working model is misspecified. Therefore, this simulation study confirms the consistency of the NPS estimator regardless the working survival model from which the prognostic tool was derived. The simulations results, however, gave the impression that the Cox proportional hazards model systematically leads to a smaller AUC compared to the MCM. This important difference is investigated in more detail in the next section.

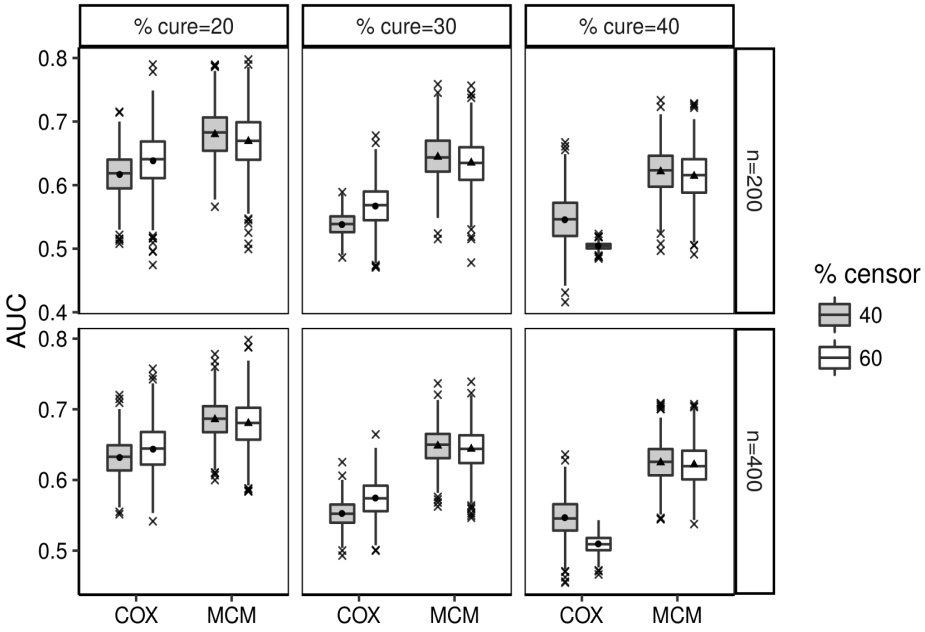

Figure 2.3: The boxplots of the averaged NPS AUC estimates $AA\hat{U}C_{\hat{M}_k^{\mathcal{W}}}$ obtained with MCM-case 2 for a time horizon $t = 2$. The solid circles and triangles indicate the true or the population version of the AUC ($AUC_{\hat{M}_k^{\mathcal{W}}}$) of the Cox and mixture cure working models, respectively.

Table 2.5: The percent bias and MSE for MCM-case 2 (Approach 2).

$t = 1$			$n = 200$			$n = 400$		
% cure	% cens	Model	$AAUC_{\hat{M}^{\mathcal{W}}}$	% bias	MSE ($\times 10^{-3}$)	$AAUC_{\hat{M}^{\mathcal{W}}}$	% bias	MSE ($\times 10^{-3}$)
20	40	MCM	0.694	0.197	1.475	0.697	0.015	0.765
20	40	COX	0.647	0.141	1.596	0.665	0.060	0.800
20	60	MCM	0.689	0.252	1.804	0.695	-0.034	0.911
20	60	COX	0.674	0.355	1.828	0.680	0.178	0.920
30	40	MCM	0.654	0.164	1.675	0.657	0.040	0.884
30	40	COX	0.558	0.072	1.824	0.581	0.003	0.945
30	60	MCM	0.649	0.117	1.887	0.654	0.026	1.022
30	60	COX	0.602	0.183	1.969	0.613	0.111	1.050
40	40	MCM	0.625	0.069	1.969	0.628	0.156	0.981
40	40	COX	0.594	0.072	1.814	0.596	0.048	0.995
40	60	MCM	0.621	-0.018	2.107	0.626	0.022	1.064
40	60	COX	0.509	-0.044	2.278	0.522	-0.074	1.194
$t = 2$								
20	40	MCM	0.680	0.216	1.638	0.686	-0.009	0.818
20	40	COX	0.617	0.107	1.758	0.631	0.014	0.857
20	60	MCM	0.669	-0.042	2.497	0.680	-0.127	1.245
20	60	COX	0.638	0.171	2.494	0.643	0.121	1.214
30	40	MCM	0.644	0.070	1.577	0.648	0.053	0.745
30	40	COX	0.538	0.161	1.692	0.553	-0.025	0.851
30	60	MCM	0.635	-0.095	2.092	0.644	-0.123	1.010
30	60	COX	0.567	0.130	2.139	0.574	0.008	1.092
40	40	MCM	0.621	0.040	1.585	0.625	0.066	0.815
40	40	COX	0.545	0.177	1.577	0.547	-0.028	0.853
40	60	MCM	0.614	-0.056	1.978	0.622	-0.144	0.989
40	60	COX	0.504	0.005	1.975	0.509	-0.065	1.039

2.3.3 Deriving risk tools from survival models that ignore the presence of cure

The objective of this section is to investigate the loss, in terms of predictive performance, when deriving risk tools from a working survival model that does not acknowledge the presence of cured subjects. This objective is motivated from the current practice of deriving risk tools in cardiovascular disease prevention and oncology by fitting standard survival models, mainly the Cox proportional hazards model, even though a large fraction of the population are non-susceptible to experience that event during their lifetime. Two criteria were used to assess the loss in predictive ability. In the first criteria, we compared the distribution of the AUC estimates of a risk tool derived from a Cox proportional hazards model (the misspecified model) to the distribution of the AUC estimate of a risk tool derived

from a mixture cure model (the true model). In both models, the derived risk tools aim at estimating $P(T \leq t|\mathbf{X})$ for a fixed time horizon t and both include exactly the same covariates. In the first criteria we aimed to show that, in average and for a given data set, a risk tool derived from a working survival model that does not ignore the presence of cure (i.e. the MCM model) produces better predictive performance in terms of the AUC estimate compared to a working survival model that ignore the presence of cure (i.e. the Cox proportional hazards model). More specifically, we want to show that:

$$\mathrm{E}(\hat{AUC}_{\hat{M}^{\mathcal{MCM}}}(t)) \geq \mathrm{E}(\hat{AUC}_{\hat{M}^{\mathcal{COX}}}(t)). \tag{2.8}$$

In the second criteria, we compared the distribution of the true AUC of a risk tool derived from a Cox proportional hazards model (the misspecified model) to the distribution of the true AUC of a risk tool derived from a mixture cure model (the true model). Here we want to show that the true AUC of a risk tool derived from a working survival model that acknowledge the presence of cure is higher than the one derived from a working survival model that ignore the presence of cure. More specifically, we aimed to show that:

$$\mathrm{E}(AUC_{\hat{M}^{\mathcal{MCM}}}(t)) \geq \mathrm{E}(AUC_{\hat{M}^{\mathcal{COX}}}(t)). \tag{2.9}$$

For both criterias, we used some of the results provided in the previous simulations and therefore no additional simulation plans were needed. The question in the first criteria as described in equation (2.8) can be directly answered from the boxplots in Figures 2.2 and 2.3. In fact, one can clearly see from these two figures that, for a fixed censoring rate, the estimated AUC of the risk tool derived from the Cox proportional hazard model is far smaller than the one derived from an MCM model. This is true in all situations. For the second criteria as described in equation (2.9), Figure 2.4 provides the boxplots of the true area under the ROC curve $AUC_{\hat{M}_k^{\mathcal{W}}}$ obtained in the previous section with MCM-case 2 and time horizon $t = 2$ but with $N = 1000$ iterations compared to 200 iterations in the previous simulations. From this figure one can clearly see that, compared to the Cox proportional hazards model, the average true AUC of the mixture cure model is always higher and also exhibits smaller variability.

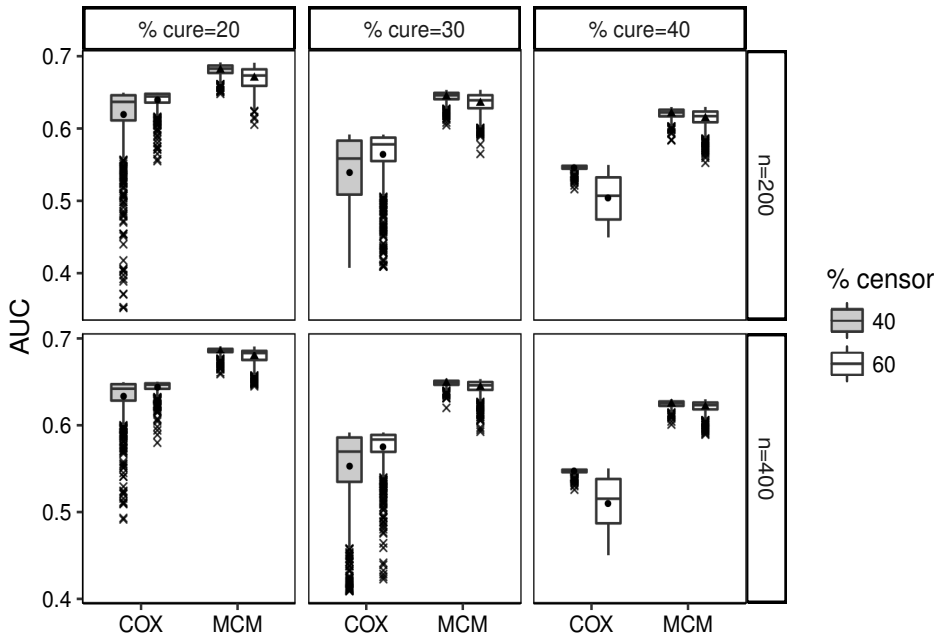

Figure 2.4: The boxplots of $AUC_{\hat{M}_k^{\mathcal{W}}}$, the population version of AUC of the prognostic tool $\hat{M}_k^{\mathcal{W}}$ obtained with MCM-case 2 for a time horizon $t = 2$ with $\mathcal{W} = MCM, COX$.

2.4 Predicting Node-Negative Breast Cancer: Should cure be taken into account?

The objective of this section is to compare, using real data example, the AUC of a survival model that do not acknowledge the presence of cured subjects with the AUC of a survival model that do acknowledge the cure. We use the data from a study on patients with lymph-node-negative breast cancer treated during $1980 - 1995$. This data set was first reported by Wang et al. (2005) and, therein, it was used to derive a prognostic tool from a Cox model. During the follow-up period, 179 out of a total of 286 patients were relapse-free (censored), and 107 of them developed distant metastasis. This is the event of interest and T, the time (measured in days) to this event, is the survival time of interest. Several factors are known to increase the risk of distant relapse. These factors include tumor size, involvement of lymph nodes and vascular invasion, tumor type and grade, hormone receptor status, etc. In this example, we consider three factors. These are: age of

the patients which range, in the analyzed data, from 26 to 83 with median of 52 years; the estrogen receptor (ER) status with levels ER− (less than 10 fmol per mg protein) and ER+ (at least 10 fmol per mg protein) and the tumour size that range from 1 to 4 with median of 1.

To explore the possible presence of cured patients, we plot the Kaplan-Meier survival curve estimate in Figure 2.5. This figure shows that the curve levels off above 0.6 with an obvious "plateau". This, together with medical evidence, clearly indicates the presence of a potential cure fraction. Modeling this data ignoring this information and considering that all patients will relapse may lead to wrong conclusion in explaining and predicting the event of interest. To investigate this, we first estimated the conditional survival distribution of T given age, ER and tumour size using (i) the standard Cox model and (ii) the MCM and then we evaluated the predictive ability of each obtained prognostic tool (survival probability) using ROC curve and AUC. We followed two approaches, in a first analysis we used all the patients as learning and validation set, whereas, in a second analysis we randomly split the 286 patients into a learning set and a validation set with 65/35 ratio. The learning set was used to estimate the prognostic tools (Cox and MCM) and the validation set to estimate the ROC and the AUC. Figure 2.6 and Figure 2.8 show the ROC curve and AUC derived according to the first approach for $t = 1$ year and $t = 3$ years, respectively. For the second approach we generate 1000 pairs of randomly selected learning and validation sets. From each pair we estimate the tools and their corresponding AUC and ROC curve. The average of these estimates for $t = 1$ year and $t = 3$ years are presented in Figure 2.7 and Figure 2.9, respectively. For the first approach, the estimated AUC values for the prognostic tool derived from MCM and Cox model are 0.694 and 0.556 for $t = 1$ year and 0.629 and 0.568 for $t = 3$ years, and for the second approach, the resulting estimates are 0.673 and 0.509 for $t = 1$ year and 0.598 and 0.516 for $t = 3$ years for MCM and Cox model, respectively. This analysis shows that the risk tool derived from the MCM has a better discrimination ability compared to the one derived from the standard Cox model as the ROC curve of the latter is clearly below the ROC of the former. This is also attested by the small AUC obtained from the Cox model. This data analysis confirms our findings in the simulation study (see Section 2.3.2).

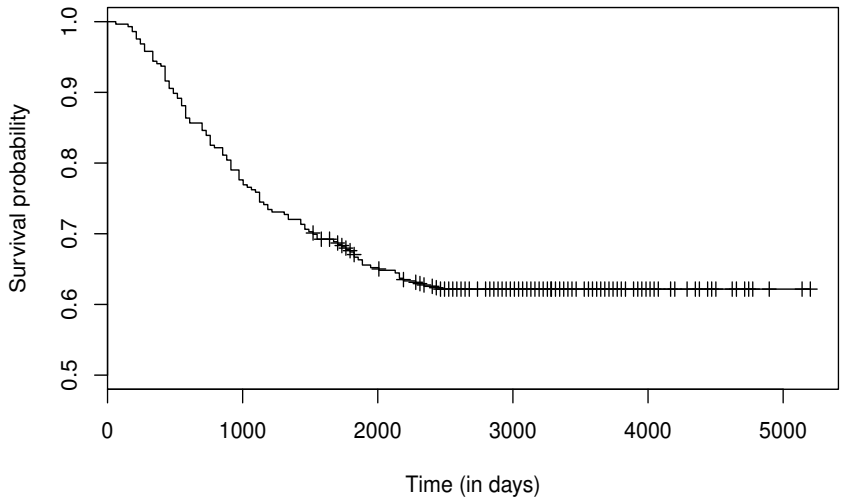

Figure 2.5: Kaplan-Meier survival estimate of the lymph-node-negative breast cancer patients (+ : censored observations).

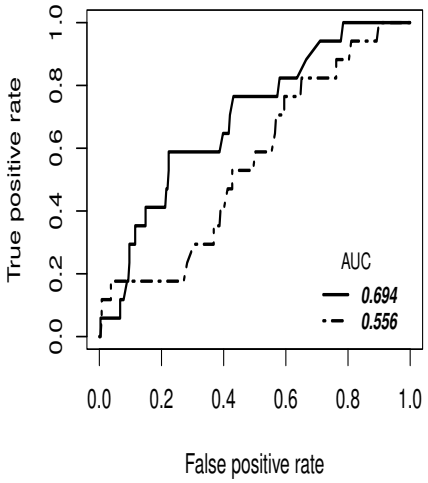

Figure 2.6: ROC curve and the corresponding AUC estimates for the MCM (solid line) and Cox model (dashed line) obtained for $t = 1$ year and using all data as training and validation set.

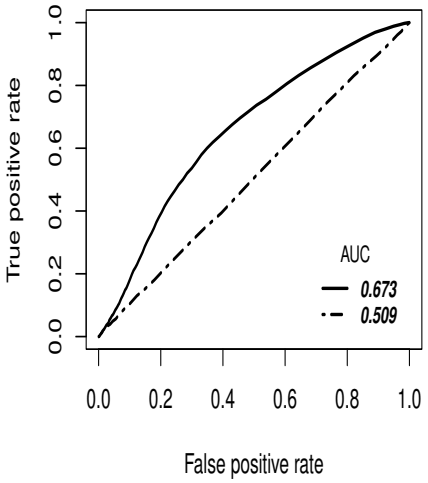

Figure 2.7: ROC curve and the corresponding AUC estimates for the MCM (solid line) and Cox model (dashed line) obtained for $t = 1$ year and by splitting the data to training and validation set.

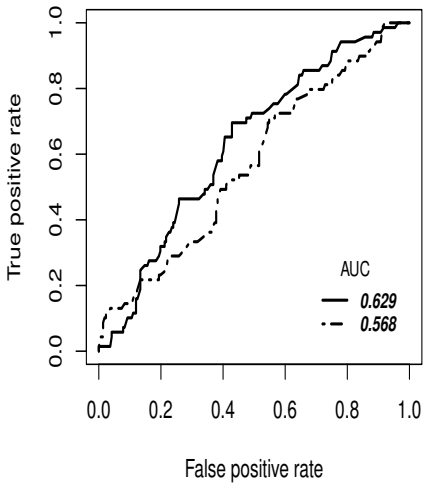

Figure 2.8: ROC curve and the corresponding AUC estimates for the MCM (solid line) and Cox model (dashed line) obtained for $t = 3$ years and using all data as training and validation set.

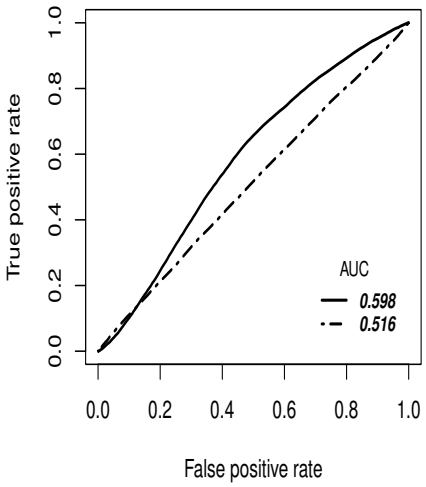

Figure 2.9: ROC curve and the corresponding AUC estimates for the MCM (solid line) and Cox model (dashed line) obtained for $t = 3$ years and by splitting the data to training and validation set.

2.5 Discussion and conclusions

In this chapter, we studied the validity of a classification accuracy measure estimator based on right censored survival data with a cured fraction. More precisely, we assessed the performance of the NPS AUC estimator by means of a comprehensive simulation with different censoring rates, cure rates, sample sizes and time horizons. We considered both the case of a known and estimated prognostic tools. In both cases, we found that the NPS method provides a consistent estimator of the true AUC for the time-to-event data with a cure fraction. This is not surprising as the method is mainly based on Beran's estimator and the latter is consistent even when there are cured subjects in the data (López-Cheda et al.; 2017). Note that, in real life, it is often difficult to tell whether the data truly contain a cure fraction or not. Therefore, having a method for estimating the time-dependent AUC that works regardless of whether the cure fraction exists or not is very useful from a practical perspective.

We also investigated the loss, in terms of predictive performance, when deriving prognostic tools from a working survival model that does not acknowledge the presence of non-susceptible subjects. This objective was mainly motivated from the

current practice of deriving risk tools in CVD research and oncology. This current practice proceeds by fitting standard survival models, mainly the Cox proportional hazards model, even though a large fraction of the population are non-susceptible to experience that event during their lifetime. We showed that a prognostic (risk) score derived from a working survival model that does not ignore the presence of cure (e.g. the mixture cure model) produces better predictive performance in terms of the ROC and AUC estimates compared to a working survival model that ignore the presence of cure (e.g. the Cox proportional hazards model).

In conclusion, in order to avoid bias and suboptimal performances, we advocate the use of cure models instead of classical survival models in deriving prognostic tools when data exhibits the presence of a cure fraction.

CHAPTER 3

Smoothed Time-Dependent ROC Curve for Right Censored Survival Data

Abstract

The prediction reliability is of primary concern in many clinical studies when the objective is to develop new predictive models or improve existing risk scores. In fact, before using a model in any clinical decision making, it is very important to check its ability to discriminate between subjects who are at risk of, for example, developing certain disease in a near future from those who will not. To that end, the time-dependent receiver operating characteristic curve (ROC) is the most commonly used method in practice. Several approaches have been proposed in the literature to estimate the ROC non-parametrically in the context of survival data. But, except one recent approach, all the existing methods provide a non-smooth ROC estimator whereas, by definition, the ROC curve is smooth. In this chapter we propose and study a new nonparametric smooth ROC estimator based on a weighted kernel smoother. More precisely, our approach relies on a well-known kernel method used to estimate cumulative distribution functions of random variables with bounded supports. We derived some asymptotic properties for the proposed estimator. As bandwidth is the main parameter to be set, we present and study different methods to appropriately select one. A simulation study is conducted, under different scenarios, to prove the consistency of the proposed method and to compare its finite sample performance with a competitor. The results show that the proposed

method performs better and appear to be quite robust to bandwidth choice. As for inference purposes, our results also reveal the good performances of a proposed nonparametric bootstrap procedure. Furthermore, we illustrate the method using a real data example.

This chapter is based on Beyene, K. M. and El Ghouch, A. (2020). Smoothed time-dependent receiver operating characteristic curve for right censored survival data. *Statistics in Medicine*, 1-24.

3.1 Introduction

The main focus of this chapter is to introduce some new nonparametric time-dependent *cumulative/dynamic* ROC curve and the corresponding AUC estimation methods for right censored time-to-event data. We also propose a bootstrap approach to estimate the variability and make inference about the ROC curve and the associated AUC. As a recall, let us first introduce relevant notations and definitions. Let T denote the time to occurrence of a certain event of interest, and M be a continuous marker measured at baseline. The event status of a subject at time horizon t is defined as $D(t) = I(T \leq t)$, with 1 indicating that the subject has the event before t (positive) and 0 otherwise (negative). Let us classify, for a given cutoff value m, a subject i as positive if its marker value is greater than m ($M_i > m$) and negative otherwise. Accordingly, the time-dependent *cumulative* true positive rate (TPR_t) and *dynamic* false positive rate (FPR_t) are defined as

$$\begin{aligned} \text{TPR}_t(m) &= P(M > m | D(t) = 1), \\ \text{FPR}_t(m) &= P(M > m | D(t) = 0), \end{aligned}$$

where $m \in (-\infty, \infty)$ is a fixed cutoff value. The corresponding time-dependent *cumulative/dynamic* ROC function (ROC_t) is

$$ROC_t(u) = \text{TPR}_t\{\text{FPR}_t^{-1}(u)\}, \tag{3.1}$$

where $u \in (0, 1)$ and $m = \text{FPR}_t^{-1}(u)$ is the threshold value such that $u = \text{FPR}_t(m)$. The ROC_t can be quantitatively summarized by the time-dependent area under the ROC curve (AUC_t) defined as

$$AUC_t = \int_0^1 ROC_t(u)du. \tag{3.2}$$

In the literature, for right censored data, several time-dependent ROC curve estimation methods that take censoring into account have been proposed. In the previous chapter, we presented a quick review of some of these methods.

Like a cumulative distribution function, the ROC is commonly estimated nonparametrically using empirical methods. This simple and very popular approach leads to a valid and reliable estimator. However, unlike the population ROC which is continuous and smooth, the empirical ROC is a step function. This is a serious drawback in many real applications where the ROC curves are used not only to evaluate and compare the discriminatory ability of markers but also to find an optimal cut-off point. To achieve this, the most commonly used methods are the point on the ROC curve closest to $(0, 1)$ and the point on the ROC curve farthest from the non-informative marker (diagonal line), also called the Youden index; see, for example, Sima and Gönen (2013); López-Ratón et al. (2014), for more details. Finding a cut-point from a smooth ROC curve is clearly more convenient as the empirical ROC will typically lead to a non-unique solution. Fluss et al. (2005) compare empirical method with a kernel based method and come to the conclusion that the former "has the worst performance and is not recommended unless sample size is very large" and advocate the use of kernel method in most situations. We shall not discuss further the problem of cut-point in the present work and refer interested readers to Hsieh and Turnbull (1996); Perkins and Schisterman (2006); Hwang et al. (2018) for more information. In their refined analysis, Lloyd and Yong (1999) compare, theoretically and via simulations, the empirical ROC estimator with a kernel estimator, in the case of uncensored data, and prove that the latter dominates the former for moderate to large samples. Similar conclusions hold when comparing the classical empirical and the kernel smooth estimator of a distribution function, see Reiss (1981); Falk (1983); Swanepoel (1988). These and other related works show that kernel based estimators are, in general, asymptotically more efficient than corresponding classical empirical estimators. Another argument in favor of smoothing is that the discontinuity of the empirical ROC makes its graphical representations not very nice from a visual point of view and not intuitive. All these justify our interest on developing a smooth nonparametric estimator for the time-dependent ROC for censored data. Note that, in contrast to cross-sectional studies, where there is an abundance literature about smoothing ROC curves, see for example the recent paper of Pulit (2016) and the references given therein, this problem is not well studied in the context of survival analysis. In fact, to the best of our knowledge, the only related work is the recent paper of Martínez-Camblor and Pardo-Fernández (2018). These authors proposed a smooth time-dependent ROC curve estimators for both *cumulative* sensitivity/*dynamic* specificity and *incident* sensitivity/*dynamic* specificity. Their estimators are based on bivariate kernel density estimation and they showed that their *cumulative/dynamic* smoothed ROC estimator outperforms over the nearest neighbor estimator of Heagerty et al. (2000). This method, however, has two main limitations. First, it has a boundary problem that produces unreliable estimates near the extremes of the ROC curve, which arises from the boundedness of both the event time and the marker. Second, there is no proposed optimal bandwidth parameter selection method.

To construct our kernel based smooth estimator we use the fact that the time-dependent ROC can be expressed as a weighted cumulative distribution of a random variable supported on the unit interval. The weights are nothing but a random variable used to impute the "missing" values of $I(T_i \leq t)$ for censored observations. In this proposal there are some important challenges that must be addressed. The first is the bounded support of ROC function. This is a problem because standard kernel estimates suffer from boundary effects when the support of the variable of interest is bounded. As this may affect the performances (convergence rate and bias) of the estimator near the boundaries, a "boundary correction" is needed. For that, one of the most used method, that we adopt in this work, is to use a transformation mapping the unit interval to the real line. After transforming the observed data, standard kernel method may be satisfactorily applied without further consideration. The second difficulty is that the weight variable is unknown and hence needs to be estimated from the observed data. For this we use a plug-in method based on Beran's nonparametric estimator (Beran; 1981). As an additional problem, in addition to the smoothing parameter needed to compute Beran's estimator, we need to choose a bandwidth parameter for our kernel estimator of the ROC curve itself. For that, we adapt some well-known existing bandwidth selection methods, used in kernel estimation of a classical (unweighted) cumulative distribution, to our situation.

The rest of this chapter is organized as follows. Next section gives the main statements and definitions of the chapter and describes, in detail, our estimation approach. In this section, we also present some asymptotic results and introduce a bootstrap procedure for estimating the standard deviation and constructing confidence intervals. Section 3.3 is devoted to the problem of the bandwidth selection. In Section 3.4, the finite sample performances of our estimator is evaluated by simulations. A real data illustration is provided in Section 3.5. Finally, discussions and conclusions are presented in Section 3.6.

3.2 Estimation method

We assume that the observed data set consists of $\{(Y_i, \Delta_i, M_i), i = 1, 2..., n\}$ where $Y_i = \min(T_i, C_i)$ is the observed survival time, T_i is the time to the event of interest, C_i is the censoring time, $\Delta_i = I(T_i \leq C_i)$ is the censoring indicator and M_i is the marker observed at the ith subject. In the following, the event status of interest will be denoted by $D(t) = I(T \leq t)$, with 1 (0) indicating that the event happened before (after) a given time horizon, t.

3.2.1 Imputation

As mentioned in the Introduction, our objective is to measure the predictive ability of M using a smooth ROC curve. The main difficulty is that event status $D(t)$ is unknown for some subjects as a result of censoring. Hence, following an idea suggested independently by Martínez-Camblor et al. (2016) and Li et al. (2018), we impute $D(t)$ by its expected value given the observed data. Thus, we impute $D(t)$ by the random variable

$$W(t) = P(T \leq t)|Y, \Delta, M) = \Big[1 - (1 - \Delta)\frac{S(t|M)}{S(Y|M)}\Big] I(Y \leq t),$$

where $S(t|m) = P(T > t|M = m)$ denotes the conditional survival function of T given the marker M. The second equality in the above equation is true under the usual assumption, in survival analysis, that T and C are independent given M. The detailed mathematical development of this weight is given in Section 2.2.1. Unlike other existing methods in the literature, here all the available information is used in the imputation. Also, it is interesting to observe that $W(t) = D(t)$ except in the case when $Y = y < t$ and $\Delta = 0$. In this case, $W(t)$ is nothing but the conditional probability $P(T \leq t|T > y, M)$. In the above expression of $W(t)$, $S(t|m)$ is unknown and needs to be estimated form the observed data. This can be done via, for example, the kernel estimator of Beran that can be written in the following form:

$$\hat{S}(t|m) = \prod_{i:\, Y_i \leq t} \left(1 - \frac{v_i(m)}{\sum_{j=1}^{n} v_j(m) I(Y_j \geq Y_i)}\right)^{\Delta_i},$$

where, for a given bandwidth $b \equiv b_n$ and a given kernel density k, $v_i(m) = \frac{k(\frac{M_i - m}{b})}{\sum_j k(\frac{M_j - m}{b})}$. The consistency and asymptotic normality of this estimator were established by several researchers (see, for example, Beran (1981); Dabrowska (1989); González-Manteiga and Cadarso-Suárez (1994); Van Keilegom and Veraverbeke (1997)). In the following, we will use the notation $\hat{W}_i(t)$ to denote $\Big[1 - (1 - \Delta_i)\frac{\hat{S}(t|M_i)}{\hat{S}(Y_i|M_i)}\Big] I(Y_i \leq t)$.

3.2.2 Estimation

In this section, we first introduce an empirical (non-smooth) estimation of the ROC and from this we propose a smooth version of it. As given by (3.1), the ROC can

be written as

$$\begin{aligned}\mathrm{ROC}_t(u) &= TPR_t\{FPR_t^{-1}(u)\} = P(FPR_t(M) \leq u | D(t) = 1) \\ &= \frac{P(Z(t) \leq u, T \leq t)}{P(T \leq t)} = \frac{E(I(Z(t) \leq u, T \leq t))}{E(I(T \leq t))} \\ &= \frac{E\{E(I(Z(t) \leq u, T \leq t)|Y, \Delta, M)\}}{E\{E(I(T \leq t)|Y, \Delta, M)\}} \\ &= \frac{E\{I(Z(t) \leq u)P(T \leq t|Y, \Delta, M)\}}{E\{P(T \leq t|Y, \Delta, M)\}} \\ &= \frac{E\{I(Z(t) \leq u)W(t)\}}{E\{W(t)\}}, \end{aligned} \tag{3.3}$$

where $Z(t)$ denotes the random variable $FPR_t(M)$, which can also be written as

$$\begin{aligned} FPR_t(m) &= \frac{P(M > m, T > t)}{P(T > t)} = \frac{E(I(M > m, T > t))}{E(I(T > t))} \\ &= \frac{E\{E(I(M > m, T > t)|Y, \Delta, M)\}}{E\{E(I(T > t)|Y, \Delta, M)\}} \\ &= \frac{E\{I(M > m)P(T > t|Y, \Delta, M)\}}{E\{P(T > t|Y, \Delta, M)\}} \\ &= \frac{E\{I(M > m)(1 - W(t))\}}{E\{(1 - W(t))\}}. \end{aligned}$$

An empirical estimator, say $\widehat{FPR}_t(m)$, of this quantity can be obtained by replacing expectations with simple averages. Plugging-in this on (3.3), leads to the following empirical estimator of the ROC,

$$\widehat{\mathrm{ROC}}_{emp}(u) = \frac{\sum_{i=1}^{n} \hat{W}_i I(\hat{Z}_i \leq u)}{\sum_{i=1}^{n} \hat{W}_i}, \tag{3.4}$$

where $\hat{Z}_i \equiv \hat{Z}_i(t) = \widehat{FPR}_t(M_i)$ and $\hat{W}_i \equiv \hat{W}_i(t)$. Here and in the following, we suppress the dependence on time t for notational simplicity.

The ROC curve estimator given in equation (3.4) is nothing but a weighted empirical distribution function with weights given by $\hat{\mathcal{W}}_i = \hat{W}_i / n^{-1} \sum_{i=1}^{n} \hat{W}_i$, i.e., $\widehat{ROC}_{emp}(u) = n^{-1} \sum_i \hat{\mathcal{W}}_i I(\hat{Z}_i \leq u)$. To the best of our knowledge, this simple and promising estimator has never been proposed or studied in the literature. However, it has a serious drawback of being a step function. For this reason, and given the arguments invoked in the Introduction, here we are interested in getting a smooth ROC curve estimator. As done to smooth out the classical empirical

distribution, see Nadaraya (1964), this goal can be easily achieved by replacing the indicator function $I(.)$ in $\widehat{\text{ROC}}_{emp}$ by a kernel distribution function. More precisely, for a bandwidth parameter $h \equiv h_n$, i.e., a sequence of strictly positive real numbers converging to zero with n, a kernel smooth estimator of $ROC(u)$ is given by

$$\begin{aligned}\widehat{\text{ROC}}(u) &= \int_{-\infty}^{\infty} k\left(\frac{u-x}{h}\right) d\widehat{\text{ROC}}_{emp}(x) \\ &= n^{-1}\sum_{i=1}^{n} \hat{\mathcal{W}}_i K\left(\frac{u-\hat{Z}_i}{h}\right), \end{aligned} \tag{3.5}$$

where $K(x) = \int_{-\infty}^{x} k(s)ds$ is a (kernel) distribution function with a density k. Clearly $\widehat{\text{ROC}}$ inherits its smoothness from K. Also, observe that when K is chosen to be the indicator function $I(x \geq 0)$ then $\widehat{\text{ROC}}$ reduces to $\widehat{\text{ROC}}_{emp}$. So that this latter can be seen as a particular case of equation (3.5).

3.2.3 Some asymptotic results

In this section we establish some asymptotic properties of the proposed estimator including its consistency and asymptotic mean squared error. For that let us assume, for simplicity, that the kernel k is a bounded symmetric (around zero) density with a bounded support, say $[-1, 1]$. Put $\mu_2(k) = \int y^2 k(y)dy$ and $\rho(k) = 2\int yk(y)K(y)dy$. The bandwidth h is such that $h \to 0$ and $nh \to \infty$. We will also need the following assumption.

Assumption A: For a given t value less than τ_m, where τ_m is the upper bound of the support of the conditional distribution function of the observed time, $H(.|m)$, such that $\inf_{m\in R_M}(1 - H(t|m)) > 0$, where R_M is the support of the marker M, (i) $\sup_m \sup_{x\leq t} |\hat{S}(x|m) - S(x|m)| = O(r_n)$ a.s., for some $r_n = o(1)$, and (ii) $\inf_m S(t|m) > s$, for some $s > 0$.

Under some suitable conditions, see Van Keilegom and Akritas (1999), Assumption $A(i)$ holds with $r_n = (\log(b^{-1})/(nb))^{1/2} = o(b^{1+\delta})$, for some $0 < \delta < 1$. This Assumption guarantees that $\max_i |\hat{W}_i - W_i| = O(r_n)$ a.s. Which, by some simple algebra, implies that $\max_i |\hat{Z}_i - Z_i| = O(r_n)$ a.s. Using these two equalities, it can be shown that, uniformly in u,

$$\widehat{\text{ROC}}(u) = ROC_n(u) + O(r_n) \text{ a.s.}, \tag{3.6}$$

where $ROC_n(u) = n^{-1}\sum_{i=1}^{n} \mathcal{W}_i K\left(\frac{u-Z_i}{h}\right)$ and $\mathcal{W}_i = \frac{W_i}{E(W_i)} = \frac{W_i}{P(T \leq t)}$.

Let f and F denote the density and cumulative distribution of Z. Define $\psi_p(z) = E(\mathcal{W}^p|Z = z)$, $p = 1, 2$, and

$$\xi_p(u) = \int_{-\infty}^{u} \psi_p(z)dF(z) = E\left\{\mathcal{W}^p I(Z \leq u)\right\}.$$

Observe that $\xi_1(u) = ROC(u)$ and

$$E\left\{\mathcal{W}^p K^p\left(\frac{u-Z}{h}\right)\right\} = E\left\{\psi_p(Z)K^p\left(\frac{u-Z}{h}\right)\right\} = \int_{-\infty}^{\infty} K^p\left(\frac{u-z}{h}\right) d\xi_p(z).$$

Standard tools from kernel estimation literature, see, for example, Chapter 1 of Li and Racine (2007), lead to the following result.

Theorem 1. *Assume that $f^{'}$, $\psi_1^{'}$ and ψ_2^2 are continuous at u, then*

(i) $E\left\{ROC_n(u)\right\} = ROC(u) + \frac{h^2}{2}\mu_2(k)ROC^{''}(u) + o(h^2)$.

(ii) $nVar\left\{ROC_n(u)\right\} = Var\left\{\mathcal{W}I(Z \leq u)\right\} - h\rho(k)\psi_2(u)f(u) + o(h)$.

By combining this Theorem and equation (3.6), we obtain the following result on the MSE of $\widehat{\mathrm{ROC}}(u)$:

$$MSE\left\{\widehat{\mathrm{ROC}}(u)\right\} = AMSE(u) + o(h^4 + \frac{h}{n} + b^{2+2\delta} + h^2 b^{1+\delta}),$$

where

$$AMSE(u) = \frac{h^4}{4}\mu_2^2(k)(ROC^{''}(u))^2 + n^{-1}Var\left\{\mathcal{W}I(Z \leq u)\right\} - \frac{h}{n}\rho(k)\psi_2(u)f(u). \tag{3.7}$$

The mean square error of $\widehat{ROC}(u)$ is of order $o(h^4 + \frac{h}{n} + b^{2+2\delta} + h^2 b^{1+\delta})$ which is different from the order of MSE of the classical (unweighted) kernel smoothed distribution estimator that is $o(h^4 + \frac{h}{n})$ (see, for example, Li and Racine (2007), p.22). The extra term $b^{2+2\delta} + h^2 b^{1+\delta}$ comes from the estimation of an unknown weight W. The order of the variance term here is $o(hn^{-1})$, which is the same as the classical kernel smoothed distribution estimator. However, the bias term has an extra term $b^{1+\delta}$ which comes from the estimation of the weight. These asymptotic results reduce to the classical case if the weights are known constant.

3.2.4 Boundary correction

As any standard kernel based nonparametric estimator $\widehat{\mathrm{ROC}}(u)$ is subject to boundary bias because $\hat{Z}$ (the pseudo data) is supported in the unit interval $[0, 1]$. This

bias comes from the fact that the kernel, which has no knowledge of the boundary, assigns positive weight outside the compact support of the variable. To be more clear, the calculation done in the previous section assumes, implicitly, that $u \in [h, 1-h]$. When, for example, $u \in [0, h)$, i.e. $u \equiv u_n = \alpha h$, for some $0 \leq \alpha < 1$, it can be shown that

$$E\left\{ROC_n(u)\right\} = ROC(u) \int_{-1}^{\alpha} k(y)dy + O(h).$$

This implies that our estimator $\widehat{\text{ROC}}$ is inconsistent in the boundary regions. Such a problem is very well documented in the literature, especially in the case of kernel density estimation, and several modified estimators have been proposed. The most used methods are the reflection method, the boundary kernel method, the local linear approach and the transformation method; see Karunamuni and Alberts (2005); Tenreiro (2013) and the references therein for more about this subject. In the case of completely observed data, Koláček and Karunamuni (2009) has recently suggested and studied a boundary correction method for the classical kernel estimator of the ROC curve. Here, the procedure we shall use to handle this problem is to apply the method of transformation that aims at "sending away" the boundaries to $\pm\infty$; see Wand et al. (1991); Marron and Ruppert (1994). Specifically, the idea consists on transforming the variable with bounded support (Z in our case) to that with an unbounded support. The distribution of the transformed variable can then be estimated using the standard kernel estimator. Let $Q : [0, 1] \to \mathbb{R}$ be any continuous and increasing function such that $\lim_{x\to 0} Q(x) = -\infty$ and $\lim_{x\to 1} Q(x) = +\infty$, since $\text{ROC}_t(u) = P(Q(Z_t) \leq Q(u)|D_t = 1)$, we suggest the following modified kernel ROC estimator

$$\widehat{\text{ROC}}_{TR}(u) = n^{-1} \sum_{i=1}^{n} \hat{\mathcal{W}}_i K\left(\frac{Q(u) - Q(\hat{Z}_i)}{h}\right). \tag{3.8}$$

The most commonly used transformations are the quantile functions of random variables with unbounded support like for example the Probit transform corresponding to the standard normal quantile function denoted hereafter by Φ^{-1}. In the case of kernel density estimation, with completely observed data, this approach was recently investigated by Geenens (2014) and Wen and Wu (2015). They show some drawbacks of this simple method and suggest some refinements. Here we stick to the "naive" approach, as given by equation (3.8), but we believe that it could be interesting to investigate this problem in more detail.

Note that when Q is chosen to be the identity function, $\widehat{\text{ROC}}_{TR}$ reduces to the estimator given in equation (3.5).

Remark. *As for the AUC, since*

$$AUC_t = \int_0^1 ROC_t(u)du = \frac{\int_0^1 E(I(Z_t \leq u)W_t)du}{E(W_t)} = 1 - E(\mathcal{W}_t Z_t),$$

an estimate of it can be obtained easily using either the empirical version of this last expression, i.e., $1-n^{-1}\sum_{i=1}^{n}\hat{\mathcal{W}}_i\hat{Z}_i$*, or by a numerical integration method, i.e.,* $\int_0^1 \widehat{ROC}_{TR}(u)du$*. The former has the advantage of being simple and independent of the bandwidth parameter needed for the smooth ROC estimator.*

3.2.5 Variance estimation and inference

To make inference about the ROC or the AUC, one needs to estimate the (asymptotic) distribution of the proposed estimators. In the following we will focus on the AUC but the same procedure can be applied to each points of the ROC curve. Given the difficulties on getting an asymptotic distribution or at least a closed and easy to estimate formula for the asymptotic variance, here we propose to use the well-known "naive" nonparametric bootstrap method of Efron (1979). This consists of first drawing, with replacement, a large number, say B, of random samples of size n from the original observed data $\{(M_i, Y_i, \Delta_i), i = 1, 2, ..., n\}$. Next, from each bootstrap sample b, $b = 1, \ldots, B$, we calculate $\widehat{AUC}_b$, the estimated bootstrap AUC statistic. The empirical variance of this bootstrap statistic, i.e.,

$$S_B^2 = B^{-1}\sum_b \left(\widehat{AUC}_b - B^{-1}\sum_b \widehat{AUC}_b\right)^2,$$

can be then used as an estimator of the variance of $\widehat{AUC}$, the original estimate of the AUC. $\{\widehat{AUC}_b, b = 1, 2, ..., B\}$ can also be used to construct asymptotic confidence intervals. For that, several approaches exist in the literature; see, for example, Efron (1981). Among them there is (i) the standard normal method: $\widehat{AUC} \pm z_{1-\alpha/2}S_B$, where z_α is the $\alpha-$quantile of a $N(0, 1)$ and (ii) the percentile method: $[\widehat{AUC}_B(\alpha/2), \widehat{AUC}_B(1-\alpha/2)]$, where $\widehat{AUC}_B(\alpha)$ is the $\alpha-$quantile of the B bootstrap sample estimates. The validity and precision of these bootstrap procedures will be investigated via simulations in Section 3.4. Note that the use of bootstrap in the context of time-dependent ROC and AUC estimations is not new but appears in several papers including Heagerty et al. (2000); Uno et al. (2007); Li et al. (2018).

3.3 Bandwidth selection

To compute the proposed estimator we have to choose the kernel and the two bandwidths: b, needed by Beran's estimator in the imputation stage, and h needed to smooth the indicator function. It is well-known from kernel smoothing literature, see, for example, Jones (1990), that the kernel function has little effect on the performance of the resulting estimates but this is not the case, in general, for the bandwidth(s) that has to be selected carefully. Now, the effect of b should be less important than the effect of h as the latter is the parameter that really

controls the degree of smoothness of our estimator. Also, observe that b intervenes only in the calculation of W_i for subjects with $Y_i < t$ and $\Delta_i = 0$. More precisely, for these subjects, we need b to get $\hat{S}(t|M_i)/\hat{S}(Y_i|M_i)$. Dabrowska (1987) showed that, uniformly in t, the bias of $\hat{S}(t|M_i)$ is asymptotically proportional to b^2. Hence, as noted by Li et al. (2018), for inappropriately larger or smaller b, the biases in $\hat{S}(t|M_i)$ and $\hat{S}(Y_i|M_i)$ are in the same direction and may cancel out to some extent, particularly when Y_i and t are close to each other. This is confirmed by our simulation study reported in the next section where we investigate the robustness of our estimator to the selection of b. So in this section we will focus on the bandwidth h that may have a significant influence on the ROC estimate.

Let us start with some theoretical considerations. The asymptotic integrated mean square error ($AMISE$) is found by integrating the $AMSE(u)$, see equation (3.7). The bandwidth which minimizes this $AMISE$ can be easily obtained by differentiation and it can be expressed as

$$h = \left(\frac{E(\mathcal{W}^2)\rho(k)}{n\mu_2^2(k)\kappa(ROC)} \right)^{1/3},$$

where $\kappa(ROC) = \int_0^1 \left\{ ROC^{''}(u) \right\}^2 du = \int_0^1 \left\{ (\psi_1(u)f(u))^{'} \right\}^2 du$, where $f(.)$ is the density of Z and $\psi_1(z) = E(\mathcal{W}|Z = z)$.

It is interesting to note that when W is a constant (non-stochastic) then $ROC(u) = P(Z \leq u)$ and h reduces to $h_F = \left(\frac{\rho(k)}{n\mu_2^2(k)\kappa(F)} \right)^{1/3}$, which is the optimal bandwidth that minimizes the $AMISE$ of the classical kernel distribution function, see, for example, Polansky and Baker (2000). The difference between h and h_F comes from $E(\mathcal{W}^2) = Var\{\mathcal{W}\} + 1$. As a consequence, when the variability of $\mathcal{W}$ increases a larger bandwidth will be needed, which sounds intuitive. $E(\mathcal{W}^2) = E\{W^2\}/(E\{W\})^2$ can be easily estimated empirically using the estimated weights $\hat{W}_i$ given at the end of Section 3.2.1.

To estimate h, the real problem comes from the estimation of $\kappa(ROC)$. In the case of the (unweighted) distribution function, the simplest and most used approach, known as Rule-of-Thumb, is to estimate $\kappa(F)$ using a parametric model. For example, if we assume that f is a normal density with mean μ and variance σ^2, then it is very simple to show that $\kappa(F) = (4\sigma^3\pi^{0.5})^{-1}$; see, for example, Polansky and Baker (2000). An estimate for $\kappa(F)$ is given by $\widehat{\kappa(F)} = (4\hat{\sigma}^3\pi^{0.5})^{-1}$, where $\hat{\sigma}$ is an estimator for the population standard deviation σ. Silverman (1986) suggests to use the minimum of sample standard deviation and sample interquartile range divided by 1.349 as a robust estimate of σ.

Since the ROC can be seen as a weighted distribution function, one may adapt the above normal reference rule and estimate $\kappa(ROC)$ by

$$\widehat{\kappa(ROC)}_{NR} = (4\hat{\sigma}_w^3 \pi^{0.5})^{-1}, \tag{3.9}$$

where $\hat{\sigma}_w$ is the minimum of weighted sample standard deviation (S_w) and weighted sample interquartile range (IQR_w) divided by 1.349. From the sample data, the quantity S_w can be computed as

$$S_w = \sqrt{\sum_{i=1}^{n} \hat{\mathcal{W}}_i (\hat{Z}_i - \bar{\hat{Z}}_w)^2},$$

where $\bar{\hat{Z}}_w = \sum_{i=1}^{n} \hat{\mathcal{W}}_i \hat{Z}_i$, $\hat{Z}_i = \widehat{FPR}_t(M_i)$ and $\hat{\mathcal{W}}_i = \hat{W}_i / n^{-1} \sum_{i=1}^{n} \hat{W}_i$.

To compute the weighted sample interquartile range, we first calculate the first and third quartiles as

$$Q_{1w} = \hat{Z}_{(c_1)} + a_1(\hat{Z}_{(c_1+1)} - \hat{Z}_{(c_1)}), \text{ and } Q_{3w} = \hat{Z}_{(c_2)} + a_2(\hat{Z}_{(c_2+1)} - \hat{Z}_{(c_2)}),$$

where the two integers c_1 and c_2 are numbers such that

$$\sum_{i=1}^{c_1} \hat{\mathcal{W}}_i \leq 0.25 \quad \text{and} \quad \sum_{i=1}^{c_2} \hat{\mathcal{W}}_i \leq 0.75,$$

and

$$a_1 = 0.25 - \sum_{i=1}^{c_1} \hat{\mathcal{W}}_i \quad \text{and} \quad a_2 = 0.75 - \sum_{i=1}^{c_2} \hat{\mathcal{W}}_i.$$

From these, the weighted sample interquartile range is

$$IQR_w = Q_{3w} - Q_{1w}.$$

The resulting bandwidth is given by

$$\hat{h}_{NR} = \left[\frac{\hat{E}\{\hat{\mathcal{W}}^2\} \rho(k)}{n \mu_2^2(k) \widehat{\kappa(ROC)}_{NR}} \right]^{1/3}, \tag{3.10}$$

and will be called *normal reference* bandwidth.

Another option is to estimate the quantity $\kappa(ROC)$ non-parametrically as it was done by Polansky and Baker (2000) for $\kappa(F)$. Following these authors, using integration by parts, we can easily show that

$$\kappa(ROC) = -E\left\{ \mathcal{W} ROC^{(3)}(Z) \right\},$$

where $ROC^{(3)}(.)$ is the third derivative of the ROC function. Which suggests estimating $\kappa(ROC)$ by

$$\kappa\widehat{(ROC)}_{PI} = -\hat{E}\Big\{\hat{\mathcal{W}}\widehat{ROC^{(3)}}(Z)\Big\} = -n^{-2}g^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{\mathcal{W}}_i\hat{\mathcal{W}}_j K^{(3)}\Big(\frac{\hat{Z}_i - \hat{Z}_j}{g}\Big), \tag{3.11}$$

where $K^{(3)}(.)$ is the third derivative of the kernel distribution function K and g is a smoothing parameter that needs to be chosen. We will refer to the resulting bandwidth, obtained by replacing $\kappa\widehat{(ROC)}_{NR}$ by $\kappa\widehat{(ROC)}_{PI}$ in equation (3.10), as the *plug-in* bandwidth and will be denoted by h_{PI}. The main difficulty of this approach is that one still needs to choose the pilot bandwidth g in order to calculate $h_{PI} \equiv h_{PI}(g)$. For the classical (unweighted) kernel distribution estimation, Polansky and Baker (2000) proposed an iterative method where g is selected based on the asymptotic mean squared error of $\widehat{\kappa(F)}$. Due to the complexity of deriving the asymptotic mean squared error for $\kappa\widehat{(ROC)}_{PI}$, in this study we use the normal reference method given in equation (3.10) to calculate the bandwidth g.

An alternative approach is to use Cross-Validation (CV). In the case of the distribution function, different approaches exist in the literature; see, for example, Sarda (1993); Altman and Léger (1995); Bowman et al. (1998). In the latter the authors proposed a modified cross-validation estimator that is asymptotically optimal and performs well in simulation studies and real data analysis. The suggested bandwidth is obtained by minimizing the function

$$CV_F(h) = \sum_{i=1}^{n}\int_{-\infty}^{\infty}\big[I(x \geq x_i) - \hat{F}^{-i}(x_i)\big]^2 dx,$$

where $\hat{F}^{-i}(x_i)$ denotes the kernel distribution estimator calculated with the bandwidth h and without the i^{th} observation. To adapt this method to our case, we suggest incorporating the weight variable into the CV function and select the bandwidth that minimizes

$$CV_{ROC}(h) = \sum_{i=1}^{n}\int_{0}^{1}\big[\hat{\mathcal{W}}_i I(u \geq \hat{Z}_i) - \widehat{\mathrm{ROC}}^{-i}(u)\big]^2 du, \tag{3.12}$$

where $\widehat{\mathrm{ROC}}^{-i}(u) = (n-1)^{-1}\sum_{j\neq i}\hat{\mathcal{W}}_j K(\frac{u-\hat{Z}_j}{h})$ is the ROC curve estimate without the i^{th} observation.

The performances of the three methods discussed above (normal reference (NR), plug-in (PI) and Cross-Validation (CV)) will be investigated in the next section.

3.4 Simulation Study

In this section, we present a Monte Carlo simulation study to investigate the finite sample performance of the proposed ROC curve estimators: the empirical (non-smoothed), see equation (3.4), the smoothed estimator without boundary correction, see equation (3.5), and the smoothed estimator with boundary correction, see equation (3.8). We are also interested in comparing the performance of the latter with the smoothed time-dependent ROC curve estimator proposed by Martínez-Camblor and Pardo-Fernández (2018), hereafter denoted by (MP). As mentioned above, the ROC curve estimator with and without the boundary correction requires selecting two bandwidth parameters. Given the difficulty to select these parameters in an optimal way, we also investigate here the robustness of these ROC estimators to the choice of the bandwidth parameters and assess the performance of the proposed modified bandwidth selection methods. Furthermore, the validity of the proposed bootstrap approach is also examined. The behavior of the estimator depends on various aspects, including sample size and censoring rates. Therefore, we generate data considering different scenarios, the details of which are given in the upcoming subsection.

3.4.1 Data generation

To generate the data, we considered two scenarios

- Scenario I: The survival time T is assumed to follow a log-normal distribution with parameters $\mu = 0$ and $\sigma = 2$. The censoring time C is independently generated from a log-normal distribution with parameters μ and $\sigma = 2$. The value of μ is chosen to produce the required censoring proportions. The marker is given by $M = \sqrt{p}T + \sqrt{1-p}R$, where R is a log-normal distributed random variable with parameters $\mu = 0$ and $\sigma = 2$.

- Scenario II: The survival time is assumed to follow an $Exponential(1)$ distribution. The censoring time is taken to be independent of T and also follows an exponential distribution with parameter λ, where the value of λ is chosen to produce the required censoring proportions. The marker is given by $M = \sqrt{p}T + \sqrt{1-p}R$, where R is an $Exponential(1)$ distributed random variable.

These scenarios are very similar to those of Martínez-Camblor and Pardo-Fernández (2018). The values of p are set to be 0.25 and 0.50. The later corresponds to a stronger association between the survival time and the marker. The parameter of the censoring time distributions are chosen to achieve 20% and 50% censoring rates. Furthermore, for the ROC curves calculation, we consider the time horizon t corresponding to the quartile values (Q_1, Q_2 and Q_3) of the survival time T. For both scenarios, we generate $N = 1000$ replicates of samples of sizes $n = 100$,

200, 400. The true time-dependent ROC curves for both scenarios are presented in Figure 3.1.

The performance measures for the proposed ROC estimator with and without boundary correction are calculated using the three proposed data-driven bandwidth selection methods for h: the normal reference bandwidth (NR), the plug-in (PI) and the cross-validation (CV). To select b, the bandwidth parameter needed for Beran's weight calculation, we simply use the plug-in bandwidth selection method of the classical density estimation that was proposed by Sheather and Jones (1991). This method is implemented in the `bandwidth` function of the `stats` package of R (R Core Team; 2020).

The MP estimator of Martínez-Camblor and Pardo-Fernández (2018) is based on the bivariate kernel density estimation and hence it uses a 2×2 matrix of smoothing parameters. In their simulations, the authors considered six data-driven bandwidths selection methods: plug-in (PI), least-squared cross-validation (LSCV) and unbiased cross-validation (UCV), based either on diagonal or non-diagonal smoothing matrices. They found that the non-diagonal version has better performance and hence in our comparison we only consider this latter. The MP method is implemented in the `R` package `smoothROCtime` (Díaz-Coto; 2018), which was used in our simulations.

To compare the performance of the ROC estimators we chose two criteria: the integrated mean bias given by

$$\text{MIB}(\widehat{\text{ROC}}) = 1000^{-1} \sum_{l=1}^{1000} \int (\widehat{\text{ROC}}_l(u) - \text{ROC}(u))du,$$

where $\widehat{\text{ROC}}_l(.)$ is the ROC estimate from the l^{th} sampled data and the integrated mean square error (MISE) given by

$$\text{MISE}(\widehat{\text{ROC}}) = 1000^{-1} \sum_{l=1}^{1000} \text{ISE}(\widehat{\text{ROC}}_l),$$

where $\text{ISE}(\widehat{\text{ROC}}_l) = \int (\widehat{\text{ROC}}_l(u) - \text{ROC}(u))^2 du$. In the above formulas, the true ROC(.) was calculated from a simulated sample (T_i, M_i), $i = 1, ..., 4 \times 10^6$, using equation (3.4) with $\hat{W}_i = I(T_i \leq t)$.

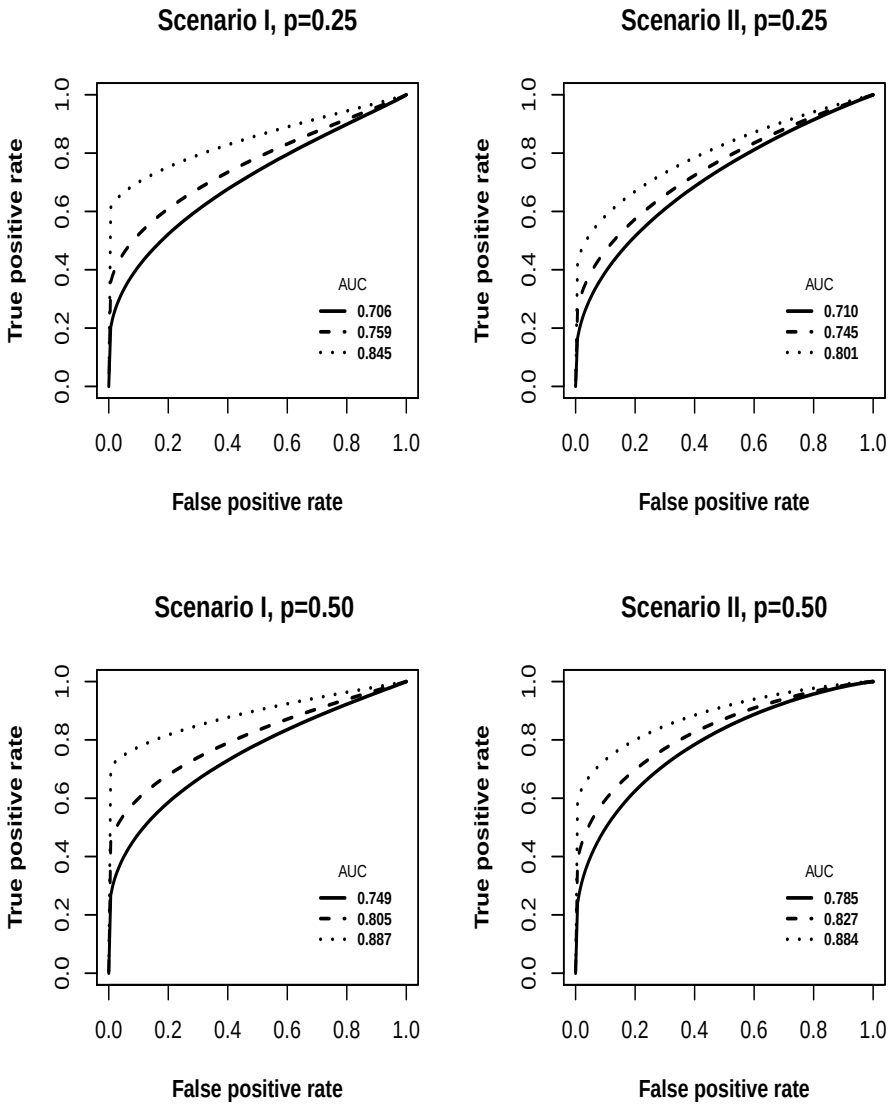

Figure 3.1: The true ROC curves and the corresponding AUC values of the first scenario (left column) and the second scenario (right column) for the prediction time Q_1 (solid line), Q_2 (dashed line) and Q_3 (doted line).

3.4.2 Simulation results

The simulation results below are organized into two parts. First, we compare the overall performance of the smooth ROC curve estimator with and without boundary correction and also compare the performance of the proposed bandwidth selection methods. Furthermore, we also explore the robustness of these estimators to the choice of the bandwidth parameters. Secondly, we compare the finite sample performances of the boundary corrected smooth ROC estimator with the empirical non-smoothed ROC estimator and the MP method of Martínez-Camblor and Pardo-Fernández (2018).

Bandwidth selection and boundary correction

We first present the results of the simulation study conducted to investigate the finite sample behavior of the ROC curve estimator with and without boundary correction given in equation (3.5) and equation (3.8), respectively. Table 3.1 shows the MIB and the MISE for both estimators for the first scenario obtained with sample sizes $n = 100$, 200, 400, censoring rates 20% and 50%, $p = 0.25$, 0.50 and $t = Q_1$, Q_2, Q_3. These results clearly show that the ROC curve estimator with boundary correction has smaller MIB and MISE than the ROC curve estimator without boundary correction. This is true for all censoring rates, sample sizes, p and t values. Both the MIB and the MISE decrease with the sample size and increases with the censoring percentages. Results for the second scenario presented in Table 3.2 also show similar findings.

Regarding the bandwidth selection methods, the results presented in Table 3.1 and Table 3.2 indicated that all the three bandwidth selection methods produce very similar results. However, the NR method has slightly better performance for the ROC curve estimator with boundary correction whereas for the ROC estimator without boundary correction the PI method seems to be better. Globally, the PI method also shows smaller bias and, typically, it performs better for a relatively large time horizon ($t = Q_3$).

Table 3.1: MIB($\times 10^{-2}$) and MISE($\times 10^{-3}$) of the ROC estimators with and without boundary correction for different sample sizes (n), censoring rates (% cen), p and t values obtained with the **first scenario** using the three bandwidth selection methods: NR, PI and CV.

			ROC with boundary correction						ROC without boundary correction					
			NR		**PI**		**CV**		**NR**		**PI**		**CV**	
p	% cen	t	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**
$n = 100$														
0.25	20	0.3	1.29	5.27	0.94	5.56	2.62	5.74	2.09	5.88	1.33	6.16	2.91	7.70
0.25	20	1.0	1.08	3.75	0.49	4.03	1.07	4.73	2.07	5.36	1.01	4.94	0.62	5.36
0.25	20	3.9	0.39	3.46	-0.06	3.19	-0.13	3.22	1.08	5.80	0.23	4.14	-0.11	3.36
0.25	50	0.3	1.82	6.38	1.51	6.69	2.68	6.68	2.44	7.30	1.77	7.57	3.07	8.75
0.25	50	1.0	2.09	6.38	1.62	6.64	1.95	6.95	2.68	8.19	1.85	7.79	1.74	8.08
0.25	50	3.9	0.51	7.49	0.20	7.19	0.07	7.36	1.02	9.84	0.39	8.24	0.03	7.49
0.50	20	0.3	1.43	5.03	0.95	5.34	3.12	5.96	2.32	6.27	1.39	6.25	2.61	8.15
0.50	20	1.0	1.06	3.51	0.42	3.61	0.57	4.23	2.02	5.77	0.93	4.74	0.25	4.17
0.50	20	3.9	0.16	2.69	-0.06	2.39	-0.11	2.39	0.44	4.05	0.05	2.90	-0.08	2.53
0.50	50	0.3	2.13	6.30	1.71	6.58	3.20	6.97	2.79	7.86	2.00	7.79	3.03	9.29
0.50	50	1.0	2.14	6.16	1.62	6.19	1.87	6.55	2.73	8.53	1.85	7.50	1.46	6.99
0.50	50	3.9	0.26	5.36	0.11	4.97	-0.01	5.04	0.55	7.11	0.21	5.70	-0.01	5.19
$n = 200$														
0.25	20	0.3	0.73	2.86	0.43	3.04	1.33	3.04	1.52	3.39	0.74	3.49	1.30	4.17
0.25	20	1.0	0.73	1.98	0.32	2.12	0.32	2.36	1.73	3.32	0.73	2.74	0.21	2.43
0.25	20	3.9	0.35	1.69	0.07	1.57	0.07	1.57	1.04	3.63	0.26	2.16	0.08	1.74
0.25	50	0.3	1.32	3.27	1.06	3.45	1.66	3.42	1.95	4.06	1.26	4.10	1.76	4.62
0.25	50	1.0	1.60	3.43	1.27	3.55	1.38	3.62	2.26	4.97	1.44	4.37	1.10	3.99
0.25	50	3.9	0.17	3.24	-0.08	3.14	-0.11	3.18	0.73	5.28	0.06	3.83	-0.10	3.36
0.50	20	0.3	0.83	2.69	0.42	2.89	1.60	3.07	1.72	3.66	0.79	3.49	0.83	4.15
0.50	20	1.0	0.72	1.84	0.28	1.89	0.20	2.04	1.66	3.69	0.65	2.64	0.20	2.09
0.50	20	3.9	0.11	1.37	-0.01	1.26	-0.01	1.25	0.34	2.28	0.03	1.48	0.02	1.41
0.50	50	0.3	1.47	3.17	1.11	3.34	1.85	3.43	2.16	4.43	1.34	4.18	1.44	4.62
0.50	50	1.0	1.53	3.32	1.15	3.32	1.21	3.39	2.16	5.34	1.32	4.24	0.99	3.57
0.50	50	3.9	0.00	2.38	-0.12	2.27	-0.14	2.28	0.22	3.58	-0.08	2.60	-0.12	2.46
$n = 400$														
0.25	20	0.3	0.47	1.39	0.22	1.47	0.64	1.50	1.14	1.84	0.44	1.76	0.37	1.98
0.25	20	1.0	0.38	0.98	0.11	1.04	0.07	1.11	1.28	2.06	0.38	1.45	0.07	1.18
0.25	20	3.9	0.08	0.89	-0.06	0.87	-0.06	0.85	0.67	2.35	0.03	1.16	-0.04	1.01
0.25	50	0.3	0.92	1.68	0.71	1.77	0.97	1.76	1.47	2.31	0.84	2.19	0.80	2.34
0.25	50	1.0	1.05	1.74	0.83	1.78	0.84	1.82	1.71	3.00	0.95	2.33	0.72	1.96
0.25	50	3.9	-0.01	1.59	-0.16	1.61	-0.16	1.61	0.52	3.21	-0.08	1.98	-0.14	1.79
0.50	20	0.3	0.52	1.31	0.18	1.40	0.58	1.52	1.30	2.06	0.45	1.76	0.07	1.73
0.50	20	1.0	0.36	0.91	0.08	0.92	0.05	0.96	1.21	2.35	0.32	1.39	0.06	1.06
0.50	20	3.9	-0.04	0.80	-0.09	0.76	-0.09	0.75	0.12	1.36	-0.09	0.85	-0.07	0.90
0.50	50	0.3	1.00	1.63	0.72	1.71	1.01	1.75	1.64	2.61	0.88	2.25	0.57	2.10
0.50	50	1.0	0.93	1.62	0.68	1.61	0.68	1.64	1.56	3.26	0.79	2.22	0.60	1.78
0.50	50	3.9	-0.11	1.23	-0.18	1.22	-0.18	1.21	0.05	1.98	-0.17	1.34	-0.15	1.38

Table 3.2: MIB($\times 10^{-2}$) and MISE($\times 10^{-3}$) of the ROC estimators with and without boundary correction for different sample sizes (n), censoring rates (% cen), p and t values obtained with the **second scenario** using the three bandwidth selection methods: NR, PI and CV.

			ROC with boundary correction						ROC without boundary correction					
			NR		**PI**		**CV**		**NR**		**PI**		**CV**	
p	**% cen**	t	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**
$n = 100$														
0.25	20	0.3	1.39	5.55	1.10	5.79	2.41	6.01	2.02	6.02	1.36	6.28	2.78	7.53
0.25	20	0.7	1.32	3.94	0.86	4.23	1.54	4.64	2.09	4.88	1.22	4.84	1.31	5.61
0.25	20	1.4	1.51	4.29	0.96	4.40	0.85	4.72	2.32	6.17	1.33	5.28	0.81	4.80
0.25	50	0.3	1.62	6.39	1.35	6.66	2.27	6.67	2.16	7.03	1.54	7.30	2.64	8.25
0.25	50	0.7	2.08	5.51	1.69	5.77	2.10	5.93	2.67	6.59	1.91	6.53	1.98	7.06
0.25	50	1.4	2.34	10.4	1.87	10.5	1.91	10.8	2.88	12.2	2.08	11.4	1.74	11.0
0.50	20	0.3	1.55	4.81	1.13	5.04	2.89	5.67	2.15	5.96	1.34	5.84	2.17	7.04
0.50	20	0.7	1.31	3.28	0.74	3.41	1.01	3.97	1.96	4.94	1.07	4.25	0.59	4.10
0.50	20	1.4	1.08	3.11	0.67	2.89	0.61	2.92	1.46	4.86	0.85	3.59	0.62	3.04
0.50	50	0.3	1.86	5.60	1.48	5.81	2.70	6.12	2.36	6.90	1.62	6.75	2.24	7.75
0.50	50	0.7	2.11	4.68	1.63	4.73	1.90	4.95	2.64	6.52	1.83	5.73	1.47	5.41
0.50	50	1.4	1.99	7.24	1.65	6.87	1.60	6.95	2.33	9.33	1.78	7.78	1.55	7.01
$n = 200$														
0.25	20	0.3	0.81	2.72	0.55	2.84	1.23	2.89	1.41	3.08	0.73	3.19	1.38	3.73
0.25	20	0.7	0.58	1.96	0.21	2.12	0.37	2.31	1.31	2.63	0.49	2.50	0.14	2.56
0.25	20	1.4	0.67	2.00	0.30	2.09	0.25	2.19	1.47	3.46	0.57	2.64	0.26	2.28
0.25	50	0.3	1.17	3.10	0.93	3.22	1.42	3.21	1.71	3.57	1.07	3.66	1.61	4.16
0.25	50	0.7	1.27	2.76	0.97	2.91	1.13	2.96	1.85	3.62	1.13	3.43	0.87	3.38
0.25	50	1.4	1.39	4.50	1.05	4.60	1.01	4.71	2.00	6.09	1.22	5.26	0.97	4.81
0.50	20	0.3	0.95	2.34	0.55	2.46	1.58	2.66	1.51	3.13	0.71	2.95	0.75	3.34
0.50	20	0.7	0.62	1.67	0.19	1.76	0.12	1.91	1.22	2.86	0.42	2.26	0.08	1.97
0.50	20	1.4	0.52	1.52	0.26	1.45	0.24	1.46	0.84	2.76	0.34	1.82	0.26	1.60
0.50	50	0.3	1.36	2.72	1.00	2.83	1.70	2.94	1.86	3.65	1.11	3.42	1.16	3.78
0.50	50	0.7	1.34	2.44	0.97	2.48	1.03	2.57	1.83	3.87	1.11	3.13	0.81	2.70
0.50	50	1.4	1.12	3.05	0.86	2.93	0.83	2.94	1.42	4.66	0.94	3.45	0.85	3.10
$n = 400$														
0.25	20	0.3	0.72	1.63	0.50	1.69	0.81	1.71	1.25	1.95	0.63	1.94	0.73	2.16
0.25	20	0.7	0.48	1.05	0.23	1.13	0.19	1.22	1.13	1.64	0.42	1.40	0.15	1.27
0.25	20	1.4	0.37	1.05	0.15	1.11	0.13	1.13	1.09	2.26	0.31	1.46	0.14	1.23
0.25	50	0.3	0.97	1.81	0.77	1.88	1.00	1.89	1.45	2.21	0.87	2.19	0.94	2.38
0.25	50	0.7	0.99	1.57	0.78	1.64	0.84	1.67	1.54	2.29	0.90	1.99	0.67	1.81
0.25	50	1.4	0.87	2.28	0.65	2.36	0.61	2.40	1.45	3.64	0.76	2.79	0.62	2.50
0.50	20	0.3	0.78	1.40	0.45	1.46	0.86	1.57	1.30	2.01	0.58	1.77	0.32	1.80
0.50	20	0.7	0.48	0.89	0.18	0.93	0.14	0.97	1.03	1.84	0.33	1.25	0.14	1.05
0.50	20	1.4	0.26	0.81	0.09	0.79	0.09	0.78	0.52	1.75	0.12	0.99	0.10	0.94
0.50	50	0.3	1.07	1.55	0.77	1.61	1.07	1.68	1.54	2.27	0.86	1.99	0.62	1.98
0.50	50	0.7	1.02	1.33	0.76	1.35	0.75	1.39	1.49	2.48	0.85	1.77	0.68	1.48
0.50	50	1.4	0.66	1.56	0.49	1.54	0.48	1.53	0.91	2.75	0.52	1.79	0.50	1.70

Now we turn to the question of robustness with respect to bandwidth selection. To investigate this problem, we reran the simulations but instead of using a data-driven bandwidth we choose the Beran's bandwidth to be $b = c_1 \times n^{-1/5}$, with $c_1 = 0.001, 0.1, 0.5, 1, 1.5, 2, 3$, and the ROC's bandwidth to be $h = c_2 \times n^{-1/3}$, with $c_2 = 1, 2, 3, 4, 5, 6$. Table 3.3 (Table 3.4) shows the MISEs obtained with the first scenario for $n = 100$ ($n = 400$). For the sake of brevity, the results for the other scenarios are not reported as they lead to the same general conclusion. In these tables, the columns show the effect of b for a fixed h and the rows show the effect of h for a fixed b. From these results, one can clearly see that both estimators are almost insensitive to the choice of b as the change in MISE is almost negligible. For example, at 20% censoring rate and $h = 100^{-1/3}$, as the b value increase by 5 folds the MISE increased by less than 0.17% for the ROC curve with boundary correction and increased by less than 2.32% for the ROC curve without boundary correction. The results also reveal that the ROC estimator without boundary correction is more sensitive to the choice of h than the ROC estimator with boundary correction. For example, at 20% censoring rate and $b = 100^{-1/5}$, doubling the value of h decreased the MISE by less than 10% for the ROC curve with boundary correction and increased the MISE by almost 75% for the ROC curve without boundary correction. In these tables, the minimum obtained value of the MISE using the grid search is indicated in bold and the minimum data-driven MISE is indicated in italic. From these results one can observe that the data-driven bandwidth selection methods of both the bandwidth b and h perform reasonably well, especially for a large sample size and a small censoring rate. For example, for the boundary corrected smooth ROC curve estimator with $n = 100$ and 20% censoring rate, the minimum MISE value using the grid search (for both b and h) is 3.247 versus 3.725 for the data-driven bandwidths. When $n = 400$, the corresponding values are 0.964 and 0.982, respectively. This is also true for 50% censoring rate but with relatively larger differences. This suggest that, although Sheather and Jones (1991) method does not take censoring into account, it results in a smooth ROC estimate close to the "optimal" one.

Table 3.3: The MISE($\times 10^{-3}$) of ROC estimators with and without boundary correction computed with different censoring rates (% cen), $\boldsymbol{n = 100}$, $p = 0.25$ and $t = Q_2 = 1$ for the first scenario and with $b = c_1 \times n^{-1/5}$ and $h = c_2 \times n^{-1/3}$. The last two columns correspond to the NR and the PI selection methods for h and the last row ($\hat{b}$) corresponds to Sheather and Jones's selection methods for b.

	ROC estimator **with** boundary correction							
$\boldsymbol{c_1/c_2}$	1	2	3	4	5	6	**NR**	**PI**
% cen=20								
0.001	4.212	3.754	3.491	3.491	3.875	4.730	3.762	4.058
0.100	4.164	3.664	3.337	**3.247**	3.515	4.238	3.631	4.011
0.500	4.171	3.701	3.400	3.338	3.638	4.398	3.687	4.032
1.000	4.165	3.718	3.436	3.398	3.730	4.527	3.723	4.037
1.500	4.157	3.723	3.454	3.432	3.786	4.609	3.739	4.033
2.000	4.147	3.720	3.461	3.453	3.823	4.665	3.743	4.025
3.000	4.133	3.717	3.475	3.490	3.887	4.758	3.750	4.014
$\hat{\boldsymbol{b}}$	4.148	3.725	3.472	3.470	3.848	4.698	*3.752*	4.029
% cen=50								
0.001	6.267	5.617	5.271	5.321	5.881	6.991	5.741	6.010
0.100	6.045	5.134	4.419	3.951	**3.858**	4.255	5.067	5.866
0.500	6.110	5.362	4.781	4.463	4.545	5.130	5.374	5.969
1.000	6.228	5.578	5.078	4.862	5.086	5.841	5.674	6.111
1.500	6.405	5.801	5.356	5.222	5.559	6.446	5.949	6.294
2.000	6.597	6.023	5.623	5.555	5.980	6.965	6.203	6.490
3.000	6.923	6.394	6.071	6.111	6.668	7.790	6.606	6.820
$\hat{\boldsymbol{b}}$	6.738	6.183	5.810	5.778	6.245	7.274	*6.379*	6.636
	ROC estimator **without** boundary correction							
% cen=20								
0.001	5.657	9.841	17.716	26.530	34.479	41.175	5.330	4.931
0.100	**5.303**	9.332	17.147	25.964	33.945	40.680	5.096	4.785
0.500	5.426	9.518	17.349	26.160	34.126	40.846	5.176	4.836
1.000	5.526	9.647	17.489	26.298	34.256	40.967	5.253	4.886
1.500	5.597	9.732	17.582	26.390	34.344	41.048	5.306	4.916
2.000	5.649	9.794	17.650	26.459	34.409	41.109	5.340	4.932
3.000	5.741	9.911	17.778	26.586	34.530	41.222	5.390	4.952
$\hat{\boldsymbol{b}}$	5.679	9.828	17.686	26.494	34.443	41.140	5.361	*4.945*
% cen=50								
0.001	7.535	11.53	19.382	28.125	35.965	42.544	7.473	7.114
0.100	**5.712**	8.708	16.152	24.893	32.901	39.697	6.622	6.643
0.500	6.470	9.853	17.384	26.076	33.994	40.697	7.026	6.874
1.000	6.969	10.468	18.050	26.733	34.615	41.272	7.374	7.133
1.500	7.424	11.002	18.627	27.303	35.153	41.771	7.706	7.390
2.000	7.827	11.458	19.113	27.781	35.603	42.188	8.001	7.627
3.000	8.502	12.229	19.929	28.579	36.352	42.880	8.471	8.011
$\hat{\boldsymbol{b}}$	8.072	11.724	19.385	28.042	35.846	42.412	8.191	*7.791*

Table 3.4: The MISE($\times 10^{-3}$) of ROC estimators with and without boundary correction computed with different censoring rates (% cen), $\boldsymbol{n = 400}$, $p = 0.25$ and $t = Q_2 = 1$ for the first scenario and with $b = c_1 \times n^{-1/5}$ and $h = c_2 \times n^{-1/3}$. The last two columns correspond to the NR and the PI selection methods for h and the last row ($\hat{b}$) corresponds to Sheather and Jones's selection methods for b.

	ROC estimator **with** boundary correction							
$\boldsymbol{c_1/c_2}$	1	2	3	4	5	6	**NR**	**PI**
% cen=20								
0.001	1.094	1.015	0.970	0.968	1.016	1.125	0.987	1.078
0.100	1.111	1.029	0.980	0.977	1.027	1.139	0.996	1.101
0.500	1.086	1.014	0.973	0.977	1.034	1.152	0.988	1.077
1.000	1.062	0.998	0.965	0.975	1.037	1.162	0.979	1.052
1.500	1.053	0.992	**0.964**	0.978	1.044	1.174	0.978	1.042
2.000	1.051	0.991	0.966	0.983	1.052	1.186	0.980	1.039
3.000	1.054	0.996	0.973	0.994	1.068	1.210	0.987	1.041
$\hat{\boldsymbol{b}}$	1.053	0.994	0.968	0.984	1.052	1.185	*0.982*	1.042
% cen=50								
0.001	1.819	1.657	1.519	1.406	1.334	1.333	1.590	1.762
0.100	1.598	1.449	1.328	1.251	1.227	1.265	1.367	1.590
0.500	1.465	1.360	1.284	**1.249**	1.265	1.345	1.318	1.452
1.000	1.491	1.410	1.361	1.350	1.391	1.504	1.392	1.469
1.500	1.660	1.587	1.548	1.550	1.609	1.751	1.580	1.634
2.000	1.877	1.804	1.768	1.777	1.851	2.017	1.801	1.847
3.000	2.317	2.243	2.209	2.228	2.324	2.528	2.244	2.283
$\hat{\boldsymbol{b}}$	1.813	1.741	1.705	1.713	1.782	1.941	*1.737*	1.785
	ROC estimator **without** boundary correction							
% cen=20								
0.001	**2.542**	4.545	7.683	12.370	17.960	23.681	1.950	1.406
0.100	2.570	4.642	7.814	12.509	18.098	23.812	1.945	1.393
0.500	2.601	4.702	7.895	12.602	18.193	23.907	1.962	1.395
1.000	2.645	4.759	7.961	12.675	18.270	23.984	1.996	1.413
1.500	2.693	4.815	8.023	12.742	18.339	24.053	2.034	1.432
2.000	2.736	4.866	8.080	12.803	18.402	24.116	2.068	1.448
3.000	2.811	4.954	8.176	12.905	18.508	24.221	2.122	1.473
$\hat{\boldsymbol{b}}$	2.723	4.850	8.062	12.784	18.382	24.096	2.058	*1.445*
% cen=50								
0.001	2.793	4.141	6.897	11.434	16.995	22.739	2.518	2.158
0.100	**2.760**	4.484	7.397	11.961	17.498	23.204	2.280	1.866
0.500	2.890	4.825	7.898	12.539	18.101	23.802	2.316	1.822
1.000	3.151	5.156	8.284	12.963	18.541	24.242	2.526	1.955
1.500	3.469	5.510	8.672	13.380	18.971	24.671	2.801	2.166
2.000	3.786	5.845	9.027	13.756	19.357	25.054	3.080	2.396
3.000	4.378	6.456	9.660	14.417	20.028	25.717	3.599	2.846
$\hat{\boldsymbol{b}}$	3.686	5.738	8.911	13.632	19.229	24.927	2.995	*2.328*

3.4.3 Comparing the proposed method with the competitors

In this section, we present the results of a simulation study conducted to compare the finite sample performance of the proposed boundary corrected ROC curve estimator with the proposed empirical (non-smoothed) estimator and the MP method of Martínez-Camblor and Pardo-Fernández (2018). Table 3.5 (Table 3.6) shows the MIB and the MISE of the first (second) scenario obtained for different sample sizes, censoring rates, p values and time horizons. As can be seen from these tables, the MP method has a notably larger MIB and MISE than both the empirical (non-smoothed) and the boundary corrected smooth estimators proposed in this work; see equation (3.4) and equation (3.8), respectively. In general this is true for all censoring rates, sample sizes and time horizons. In addition, compared to the MP method, our smooth estimator has relatively faster rate of convergence as its MISE decays faster. For example, at 20% censoring rate, $p = 0.25$ and $t = 0.3$, as the sample size increases from $n = 100$ to $n = 200$ the MISE decreases by around 50% for the smooth ROC estimator with boundary correction and decreases only by around 20% for the MP method. Furthermore, we also observed that the variance of the integrated squared errors of the MP method is larger than both the proposed smooth and the empirical (non-smoothed) estimators (the results are not reported here) which is consistent with the results presented in Table 3.5 and Table 3.6. Finally, these tables show that the proposed boundary corrected ROC estimator has smaller MISE than the non-smoothed (empirical) one except for a few cases where the prediction time is relatively large (i.e., $t = Q_3$). In such a situation, the plug-in bandwidth method (PI) seems to be more adapted as it performs better than the NR method (see Tables 3.1 and 3.2) and gives results that are very close to the empirical estimator. This is consistent with the finding of a simulation study by Lloyd and Yong (1999) that was conducted to compare the empirical ROC estimator with a kernel based estimator in the case of uncensored data.

In conclusion, we can say that the proposed boundary corrected smoothed time-dependent ROC estimator has better performance than both the proposed empirical (non-smoothed) estimator and the MP method.

Table 3.5: The MIB($\times 10^{-2}$) and MISE($\times 10^{-3}$) of the empirical, the smooth (boundary corrected), and the MP estimators computed for different sample sizes (n), censoring rates (% cen), p and t values obtained with the **first scenario**. The bandwidth(s) are selected using the NR method for our estimator and the LSCV and the PI methods for the MP estimator.

			Empirical unsmoothed		**Proposed**		**MP**			
					NR		**LSCV**		**PI**	
p	**% cen**	t	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**
$n = 100$										
0.25	20	0.3	0.197	7.454	1.294	5.271	3.337	13.288	4.656	10.855
0.25	20	1.0	0.184	4.674	1.080	3.752	3.963	13.073	8.872	21.811
0.25	20	3.9	-0.076	3.141	0.389	3.461	4.464	13.437	11.02	35.811
0.25	50	0.3	0.817	8.654	1.825	6.384	2.995	9.5497	7.773	15.740
0.25	50	1.0	1.238	7.395	2.086	6.379	5.074	12.281	11.14	28.206
0.25	50	3.9	0.062	7.271	0.509	7.486	9.049	22.848	15.12	51.266
0.50	20	0.3	0.202	6.863	1.425	5.028	3.627	14.964	4.722	12.394
0.50	20	1.0	0.198	3.925	1.064	3.513	3.914	13.506	9.203	25.066
0.50	20	3.9	-0.025	2.196	0.156	2.691	4.330	11.962	10.48	35.452
0.50	50	0.3	0.989	8.154	2.126	6.304	3.958	11.793	9.025	20.339
0.50	50	1.0	1.295	6.591	2.136	6.158	6.111	14.837	12.48	35.521
0.50	50	3.9	0.041	4.857	0.259	5.359	9.904	25.819	15.58	55.915
$n = 200$										
0.25	20	0.3	0.026	3.837	0.730	2.860	3.860	10.555	5.098	9.8133
0.25	20	1.0	0.214	2.332	0.728	1.976	4.408	11.576	9.052	21.026
0.25	20	3.9	0.123	1.481	0.354	1.694	4.901	12.906	10.45	33.018
0.25	50	0.3	0.640	4.277	1.323	3.268	3.243	8.0379	7.255	13.896
0.25	50	1.0	1.058	3.874	1.599	3.431	5.112	11.524	10.11	25.000
0.25	50	3.9	-0.068	3.116	0.169	3.235	8.447	21.377	13.32	44.034
0.50	20	0.3	0.050	3.525	0.835	2.692	4.522	12.717	5.324	11.686
0.50	20	1.0	0.224	1.975	0.724	1.840	4.898	13.554	9.502	24.762
0.50	20	3.9	0.075	1.040	0.115	1.366	5.001	13.317	9.973	32.953
0.50	50	0.3	0.693	4.004	1.466	3.168	4.163	10.069	8.588	18.643
0.50	50	1.0	0.983	3.489	1.526	3.324	6.158	14.420	11.50	32.253
0.50	50	3.9	-0.070	2.108	-0.004	2.384	9.286	24.232	13.85	48.584
$n = 400$										
0.25	20	0.3	0.008	1.766	0.468	1.392	4.678	10.024	5.377	9.5197
0.25	20	1.0	0.091	1.103	0.384	0.982	5.655	13.474	9.018	20.559
0.25	20	3.9	0.003	0.729	0.079	0.894	5.834	16.270	9.859	30.817
0.25	50	0.3	0.466	2.104	0.922	1.677	3.925	8.3241	6.873	12.733
0.25	50	1.0	0.716	1.911	1.051	1.737	5.875	13.357	9.418	22.864
0.25	50	3.9	-0.102	1.523	-0.014	1.594	8.649	23.373	12.08	38.910
0.50	20	0.3	0.007	1.612	0.517	1.315	5.456	14.057	5.768	11.603
0.50	20	1.0	0.087	0.927	0.363	0.907	6.309	17.167	9.596	24.661
0.50	20	3.9	-0.003	0.506	-0.037	0.796	5.898	16.969	9.442	30.803
0.50	50	0.3	0.488	1.970	1.004	1.628	5.098	11.376	8.244	17.489
0.50	50	1.0	0.601	1.674	0.934	1.624	7.062	17.313	10.82	29.878
0.50	50	3.9	-0.093	1.001	-0.110	1.233	9.438	26.390	12.63	43.121

Table 3.6: The MIB($\times 10^{-2}$) and MISE($\times 10^{-3}$) of the empirical, the smooth (boundary corrected), and the MP estimators computed for different sample sizes (n), censoring rates (% cen), p and t values obtained with the **second scenario**. The bandwidth(s) are selected using the NR method for our estimator and the LSCV and the PI methods for the MP estimator.

			Empirical unsmoothed		**Proposed**		**MP**			
					NR		**LSCV**		**PI**	
p	**% cen**	t	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**
$n = 100$										
0.25	20	0.3	0.281	7.788	1.392	5.548	0.915	4.969	1.141	3.533
0.25	20	0.7	0.470	5.064	1.321	3.941	2.503	4.541	3.000	4.593
0.25	20	1.4	0.815	4.723	1.515	4.288	4.470	7.180	5.014	7.953
0.25	50	0.3	0.588	8.632	1.619	6.393	1.902	5.926	2.067	4.977
0.25	50	0.7	1.292	6.566	2.081	5.511	4.247	7.316	4.404	7.134
0.25	50	1.4	1.699	10.76	2.342	10.44	7.931	16.13	8.097	16.25
0.50	20	0.3	0.292	6.458	1.550	4.813	0.846	4.019	1.057	2.833
0.50	20	0.7	0.467	3.802	1.314	3.278	2.780	4.092	3.324	4.468
0.50	20	1.4	0.660	2.859	1.077	3.114	4.907	7.956	5.400	8.874
0.50	50	0.3	0.678	7.169	1.859	5.597	2.730	5.841	3.024	5.094
0.50	50	0.7	1.311	5.073	2.113	4.681	5.686	9.075	5.910	9.159
0.50	50	1.4	1.567	6.778	1.990	7.238	9.580	20.25	9.786	20.72
$n = 200$										
0.25	20	0.3	0.068	3.641	0.810	2.721	0.623	2.506	0.837	2.054
0.25	20	0.7	0.041	2.439	0.576	1.961	1.788	2.438	2.294	2.776
0.25	20	1.4	0.276	2.199	0.668	1.998	3.602	4.450	4.125	5.341
0.25	50	0.3	0.462	4.025	1.172	3.096	1.159	3.139	1.407	2.896
0.25	50	0.7	0.753	3.243	1.272	2.760	3.301	4.343	3.533	4.546
0.25	50	1.4	0.984	4.721	1.391	4.503	6.484	10.49	6.720	10.99
0.50	20	0.3	0.086	3.033	0.946	2.345	0.655	2.044	0.857	1.693
0.50	20	0.7	0.093	1.899	0.620	1.669	2.189	2.322	2.704	2.863
0.50	20	1.4	0.295	1.376	0.518	1.523	4.225	5.683	4.680	6.619
0.50	50	0.3	0.529	3.388	1.356	2.724	1.892	3.130	2.333	3.174
0.50	50	0.7	0.814	2.623	1.335	2.438	4.680	5.825	5.040	6.476
0.50	50	1.4	0.876	2.868	1.121	3.052	8.067	14.31	8.376	15.29
$n = 400$										
0.25	20	0.3	0.230	2.027	0.719	1.627	0.756	1.480	0.916	1.333
0.25	20	0.7	0.160	1.247	0.485	1.053	1.676	1.542	2.136	1.944
0.25	20	1.4	0.161	1.122	0.373	1.050	3.258	3.327	3.723	4.127
0.25	50	0.3	0.500	2.230	0.968	1.808	0.717	1.924	0.980	1.870
0.25	50	0.7	0.672	1.775	0.995	1.567	2.657	2.912	2.921	3.211
0.25	50	1.4	0.635	2.389	0.867	2.276	5.257	7.301	5.523	7.842
0.50	20	0.3	0.209	1.689	0.784	1.401	0.839	1.211	0.972	1.110
0.50	20	0.7	0.159	0.961	0.479	0.887	2.120	1.605	2.557	2.124
0.50	20	1.4	0.142	0.690	0.256	0.812	3.987	4.897	4.357	5.657
0.50	50	0.3	0.514	1.850	1.068	1.554	1.465	1.924	1.878	2.117
0.50	50	0.7	0.690	1.403	1.019	1.328	4.094	4.310	4.448	4.998
0.50	50	1.4	0.531	1.455	0.663	1.559	6.986	11.13	7.283	12.03

3.4.4 Validity of the bootstrap

In this section, we investigate the validity of the bootstrap procedures described in Section 3.2.5 to estimate the variance and construct confidence intervals for the time-dependent AUC. For that, form each of the 1000 simulated data, we draw $B = 2000$ bootstrap samples and then calculate the bootstrap variance S_B^2 and the two confidence intervals (standard normal and percentile) using the formulas provided in Section 3.2.5. The consistency of S_B will be assessed by comparing its average (here after ASD_B) with the simulation-based empirical standard-deviation $ESD = \sqrt{1000^{-1} \sum_l \left(\widehat{AUC}_l - 1000^{-1} \sum_l \widehat{AUC}_l \right)^2}$, where $\widehat{AUC}_l$ is the AUC estimate from the lth simulated data. On the other hand, the validity and precision of the proposed confidence intervals will be assessed by calculating their coverage probabilities (CP) and their average widths (AW), respectively. The former refers to the proportion of times that the confidence interval encloses the true AUC value and the later is the average difference between the lower and upper limits of the confidence intervals. Table 3.7 shows the obtained results for data generated according to the first scenario with sample sizes $n = 100, 400$, censoring rates 20% and 50%, $p = 0.25, 0.5$ and $t = 0.3, 1.0, 3.9$. From this table, we can first notice that ASD_B increases with the censoring rate and decrease with the sample size. Also, ASD_B remains close to ESD and the difference between these two quantities decreases as sample size increases. Next, regarding the confidence intervals, we notice that, globally, for all the cases, the coverage probabilities are close to the nominal level of 95% and improve (deteriorate) as n (% of censoring) increases. On the other hand, the average widths decrease (increase) with sample size (censoring). The percentile method leads to the best performances, both in terms of coverage probability and average width. The classical normal method gives some unsatisfactory results when the sample size is small and/or the percentage of censoring is high. These results clearly demonstrate that the proposed bootstrap provides a good approximation of the sampling distribution of $\widehat{AUC}$.

Table 3.7: The empirical standard deviation (ESD), the average bootstrap standard deviation (ASD_B), the coverage probability (CP) and the average width (AW) of 95%−bootstrap confidence intervals obtained with the smooth estimator using the standard normal method (Normal) and the percentile method computed for different sample sizes (n), censoring rates (% cen), p and t values. Data simulated according to the **first scenario.**

					Normal		**Percentile**	
p	**% cen**	t	ESD	ASD_B	**CP**	**AW**	**CP**	**AW**
$n = 100$								
0.25	20	0.3	0.061	0.062	0.943	0.243	0.945	0.243
0.25	20	1.0	0.050	0.050	0.942	0.194	0.945	0.193
0.25	20	3.9	0.044	0.043	0.937	0.168	0.946	0.166
0.25	50	0.3	0.067	0.066	0.936	0.261	0.943	0.260
0.25	50	1.0	0.063	0.059	0.940	0.233	0.944	0.232
0.25	50	3.9	0.066	0.061	0.890	0.240	0.930	0.236
0.50	20	0.3	0.059	0.059	0.944	0.233	0.945	0.232
0.50	20	1.0	0.047	0.046	0.934	0.180	0.935	0.179
0.50	20	3.9	0.037	0.036	0.925	0.140	0.939	0.139
0.50	50	0.3	0.065	0.064	0.933	0.249	0.932	0.248
0.50	50	1.0	0.058	0.055	0.928	0.216	0.937	0.214
0.50	50	3.9	0.054	0.050	0.890	0.197	0.917	0.194
$n = 400$								
0.25	20	0.3	0.030	0.031	0.956	0.122	0.959	0.122
0.25	20	1.0	0.024	0.025	0.956	0.096	0.952	0.096
0.25	20	3.9	0.020	0.020	0.940	0.079	0.948	0.079
0.25	50	0.3	0.033	0.034	0.956	0.131	0.955	0.131
0.25	50	1.0	0.029	0.029	0.943	0.115	0.954	0.115
0.25	50	3.9	0.029	0.028	0.928	0.109	0.944	0.108
0.50	20	0.3	0.029	0.030	0.954	0.116	0.958	0.116
0.50	20	1.0	0.022	0.023	0.957	0.089	0.955	0.088
0.50	20	3.9	0.017	0.017	0.939	0.067	0.947	0.066
0.50	50	0.3	0.031	0.032	0.954	0.125	0.953	0.125
0.50	50	1.0	0.027	0.027	0.942	0.105	0.948	0.105
0.50	50	3.9	0.024	0.023	0.922	0.090	0.945	0.089

3.5 Real data analysis

To illustrate our method, we apply it to a real data set from a randomized placebo-controlled trial of the drug D-penicillamine for treating primary biliary cirrhosis (PBC) of the liver conducted at the Mayo Clinic between 1974 and 1984. A total of 312 subjects were randomized to the treatments and among these 125 died by the end of the follow-up and the remaining are censored subjects. Heagerty and Zheng (2005) used this data to illustrate their proposed time-dependent ROC curve estimator. They fitted a Cox model and developed two prognostic scores. This data is available in the `R` package `survivalROC` (Heagerty and Saha-Chaudhuri; 2013). For the purpose of demonstration, we use here the prognostic score derived by Heagerty and Zheng using the following five variables: log(bilirubin), albumin, log(prothrombin time), edema, and age. Figure 3.2 displays the time-dependent ROC curves estimated using the proposed boundary corrected smooth ROC curve estimator, the MP method, and the proposed empirical (non-smoothed) method evaluated at the prediction times $t = 3$, 6 years. The boundary corrected smooth ROC estimator was calculated using the NR bandwidth selection method. As for the MP method, we used the non-diagonal least squares cross-validation method. The estimated ROC curves for the boundary corrected smooth and the empirical (non-smoothed) methods are very close to each other, whereas the MP method results in a smaller ROC, especially at 6 years prediction time. Table 3.8 shows the estimated time-dependent AUC with the corresponding 95% bootstrap confidence intervals for the proposed smoothed and empirical methods. The reported confidence intervals were calculated based on 2000 bootstrap samples using either the standard normal approach or the percentile method. Regarding the importance of the marker's discrimination ability, the obtained confidence intervals reveals similar decision about the significance of the AUC values, since the "null" value (i.e. 0.5) is not contained in the intervals.

As for optimal cut-point estimation, we consider the Youden index (J) which is defined as the maximum difference between the true positive rate and false-positive rate overall possible cut-off points. The cut-off point associated with this index is the method most used in practice; see Liu (2012) for more details about this subject. To calculate (numerically) the optimal cut-off point of the smooth and non-smooth estimators, we followed the approaches described in Ruopp et al. (2008). The details about these approaches can be found in Section 4.2.2. Table 3.9 shows the estimated Youden index and the corresponding cut-off point using the empirical and smoothed estimators at $t = 3$, 6 years. The estimated Youden indexes and the corresponding cut-off points for the smoothed and non-smoothed approaches are very similar, which are consistent with the estimated ROC curves.

Table 3.8: The estimated AUC value with 95% bootstrap confidence intervals using the percentile method and the standard normal method (Normal) from both the proposed smoothed and empirical (Non-smoothed) methods for $t = 3$, 6 years.

Proposed smoothed		**95% bootstrap confidence interval**	
t	**Estimate**	**Percentile**	**Normal**
3	0.892	[0.839, 0.932]	[0.844, 0.939]
6	0.873	[0.821, 0.919]	[0.824, 0.923]
Empirical (Non-smoothed)			
3	0.897	[0.847, 0.940]	[0.850, 0.944]
6	0.877	[0.826, 0.923]	[0.827, 0.926]

Table 3.9: The estimated Youden index and optimal cut-off point obtained from the smoothed and non-smoothed estimators for $t = 3$, 6 years.

	Non-smooth		**Smooth**	
t	**Youden index**	**cut-off point**	**Youden index**	**cut-off point**
3	0.708	6.691	0.660	6.651
6	0.647	6.226	0.627	6.265

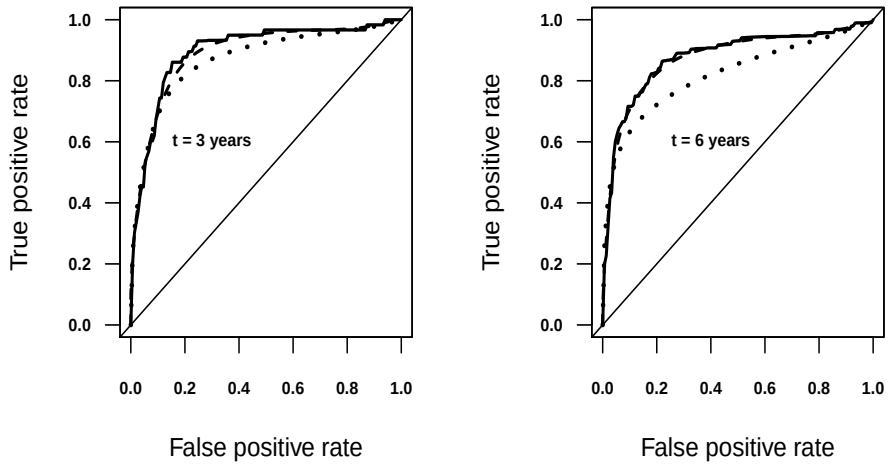

Figure 3.2: Estimated time-dependent ROC curve from the empirical (solid line), boundary corrected smooth (dashed line) and the MP method (doted line) for 3 years (left) and 6 years (right) prediction time.

In the analysis of the Mayo PBC data, the advantage of smoothing the ROC curve is not clear as the estimated ROC curves and the corresponding Youden indexes and optimal cut-off points from the smoothed and non-smoothed estimators are very close. As a result, to show the advantage of smoothing, we look into two directions detailed below: (1) showing a real data example where the two methods diverge from each other, and (2) illustrating the practical implications of a miss-estimated cut-off point and strengthening the reason why the smoothed version should be used by clinicians.

To clearly illustrate the difference between the smoothed and the non-smoothed method, we use here the classical ovarian cancer survival data from the R-package 'survival'; see Edmonson et al. (1979) for more details about this data. For the purpose of demonstration, we used the variable "age" as a marker. Figure 3.3 displays the smoothed and non-smoothed time-dependent ROC curve together with the estimated Youden index and associated optimal cut-off point estimated at $t = 370$. From this figure, one can see that the non-smoothed method leads to a larger Youden index (0.80 vs 0.75). More importantly, the optimal cut-off point estimator based on the smoothed (non-smoothed) ROC curve is about 55 (60) resulting in 5 years difference which may not be negligible in practice.

To further clarify why the smooth version of the cut-off point estimate should be use in practice, we considered a simulated data generated according to the first scenario of the simulation design given in Section 4.1 with $n = 100$, 20% of censoring, $t = 0.3$ and $p = 0.25$. For each method (smooth and non-smooth), we estimate the Youden index and the optimal cut-off point and then we compute the sensitivity and specificity at that point. Table 3.10 shows the obtained estimates as well as the corresponding true values. One can see that, in absolute value, the smoothed version of the optimal cut-off point estimate is lower than the non-smoothed one. The observed difference, here more than 4 units, may have serious consequences from the medical decision point of view. In fact, the obtained estimates based on the empirical (non-smooth) method indicates that our marker has a high sensitivity (0.96) and a small specificity (0.38) which contradict both the estimates based on smooth method (sensitivity = 0.44, specificity = 0.83) and the true sensitivity (0.53) and the true specificity (0.79) at the true optimal cut-off point. In other words, the non-smooth method largely overestimate the ability of the marker to detect a disease in patients in whom the disease is truly present and, in the same time, largely underestimate the ability of the marker to rule out the disease in patients in whom the disease is truly absent.

From all the above results we can conclude that smoothing may have a significant and positive impact on the resulting cut-off point estimate and may help to make more accurate decisions in clinical studies. For that reason, we highly recommended its use.

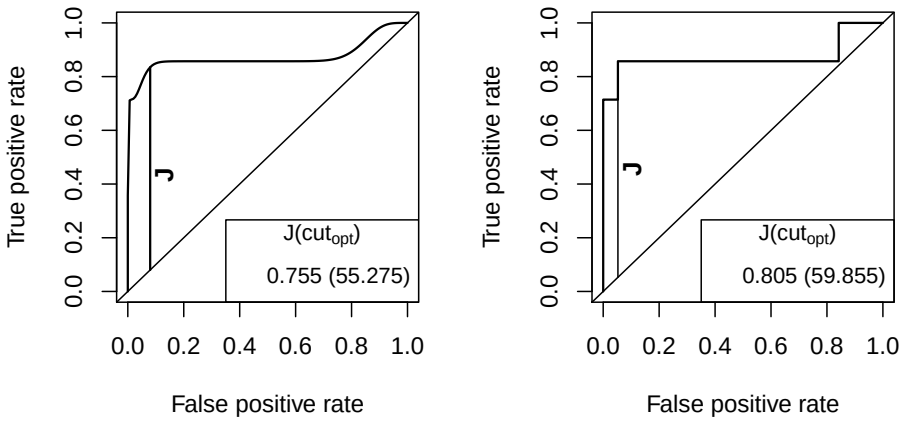

Figure 3.3: The estimated time-dependent ROC curve together with the estimated Youden index and the associated optimal cut-off point at $t = 370$ with (left) and without (right) smoothing for the ovarian cancer survival data.

Table 3.10: The true and estimated Youden index, optimal cut-off point, sensitivity and specificity obtained from a data simulated according to the **first scenario** ($t = 0.3$ and $p = 0.25$).

	Youden index	**cut-off point**	**Sensitivity**	**Specificity**
TRUE	0.315	-0.725	0.530	0.786
Smooth	0.271	-1.267	0.445	0.827
Non-smooth	0.332	-5.690	0.956	0.376

3.6 Discussion

In this chapter, we proposed and investigated a nonparametric smoothed time-dependent ROC curve estimators based on a weighted kernel smoothing technique for right censored time-to-event data. Our method relies on well-known kernel techniques used to estimate cumulative distribution functions of random variables with bounded support. We derived some asymptotic properties of the proposed estimators and suggested a boundary correction based on a quantile transformation method. An important problem in conducting kernel based estimation is the choice of bandwidth. The proposed smooth estimators require selecting two bandwidth parameters: one for Beran's weight calculation and one for smoothing the ROC curve itself. To select the latter, we proposed three data-driven methods: the

normal reference (NR), the plug-in (PI) and the cross-validation (CV). In order to estimate the variability and construct confidence intervals for the time-dependent ROC curve and its corresponding AUC, we proposed a bootstrap method.

Our simulation results show better performance, in terms of the MIB and MISE, for the boundary corrected smooth ROC estimator compared to the one without the boundary correction. Results for comparing the robustness of these estimators suggest that the latter is relatively highly sensitive to the choice of the bandwidth parameter needed for smoothing the ROC curve. However, both these estimators are almost insensitive to the choice the bandwidth needed for Beran's weight calculation. We also compared the performance of the three proposed bandwidth selection methods. From the results, we found that these methods produce very similar results with slightly better performances of the NR (PI) method for the ROC estimator with (without) boundary correction. The simulation results also show that the proposed smooth estimators lead, in general, to better results compared to the method of Martínez-Camblor and Pardo-Fernández (2018), which is, to the best of our knowledge, the only available competitor in the literature. Regarding the performance of the proposed bootstrap confidence intervals estimation approach, we found that, globally, for all the cases, the coverage probabilities are close to the nominal level of 95%. The percentile method lead to the best performances, both in terms of coverage probability and average width, compared to the classical normal method which gives some unsatisfactory results when the sample size is small and/or the percentage of censoring is high. The results clearly demonstrate that the proposed bootstrap provides a good approximation of the sampling distribution of $\widehat{AUC}$. The methods also illustrated with a real data analysis from the PBC dataset. Finally, it is important to note that, in practice, a marker is typically built from the observed data using a given statistical model (Cox, AFT, ...). In such a situation, before the model/marker can be really used, the discrimination analysis, based on the ROC or the AUC, should always be coupled with a good calibration algorithm (Cook; 2008; Van Calster et al.; 2019).

In this chapter we only considered a quantile transformation technique to correct the boundary problem. However, considering other boundary correcting approaches, such as the reflection method, and comparing its performance with the one used in this chapter is an interesting open problem. Furthermore, selecting the pilot bandwidth g of the plug-in method, see equation (3.11), based on the asymptotic mean square error is also another interesting research topic. Finally, extending and investigating the proposed smoothing method to other metrics, such as standardized net benefit and net reclassification improvement at the event rate, would be an interesting future research topic. Finally, the proposed method is implemented in the open-source `R` package `cenROC`, which is publicly available from the GitHub repository (`https://github.com/elghouch/cenROC`).

CHAPTER 4

Time-Dependent ROC Curve Estimation for Interval Censored Data

Abstract

The receiver operating characteristic (ROC) curve is the most popular graphical method for evaluating the classification accuracy of a diagnostic marker. In time-to-event studies, the subject's event status is time-dependent and hence time-dependent extensions of ROC curve have been proposed. However, in practice, the calculation of this curve is not straightforward due to the presence of censoring that may be of different types. Existing methods focus on the more standard and simple case of right censoring and neglect the general case of mixed interval censored data that may involve left, right, and interval censored observations. In this context, we propose and study a new time-dependent ROC curve estimator. We also consider some summary accuracy measures (AUC and Youden index) traditionally associated with ROC. The proposed method uses available data very efficiently. To this end, the unknown status (positive or negative) of censored subjects, are estimated from the data via the estimation of the conditional survival function given the marker. For that, we investigate both model-based and nonparametric approaches. We also provide variance estimate and confidence intervals using Bootstrap. A simulation study is conducted to investigate the finite sample behavior of the proposed methods and to compare their performance with a competitor. Globally, we observed better finite sample performances for the proposed estimators. Finally,

we illustrate the methods using two data sets one from hypobaric decompression sickness study and the other from an oral health study. The proposed methods are implemented in the `R` *package* `cenROC`.

This chapter is based on Beyene, K. M. and El Ghouch, A. (2020). Time-dependent ROC curve estimation for interval censored data, [*submitted*].

4.1 Introduction

In this chapter, our focus once again is on introducing some time-dependent ROC curves and the associated summary measures estimation and inference methods. As the summary measure for the ROC curve, we consider the two popular indices: the area under the ROC curve (AUC) and the Youden index. Unlike the preceding two chapters which focused on the right censored data, here, we consider the general case of mixed interval censored data that may involve left, right, and interval censored observations. This is mainly due to the fact that, although many practical situations give rise to interval censoring, literature about the time-dependent ROC curve and the corresponding summary measures estimation methods that take the interval censored nature into account are scarce as compared to the right censored data. One practical example that gives rise to interval censoring is when the subjects are assessed periodically for the response of interest. As a consequence, the time when the event of interest occurs is not exactly observed but only known to have occurred at some time between two consecutive visits. The time of occurrence of this event is said to be interval censored. Compared to the right-censored data, the literature about ROC and associated summary measures that take this phenomenon into account are scarce. The shortage of literature, in this case, may be explained by the technical difficulties that arise due to a relatively severe loss of information so that fundamental quantities, like the survival function, that can easily be computed under, for example, right-censoring, do not have a closed-form formula under interval-censoring. An ad hoc approach to deal with interval-censored observations is to impute each interval by its mid-point or right-point and then use classical methods proposed for uncensored or right-censored data. Unfortunately, this naive approach is known to result in biased estimates; see, for example, Sun and Chen (2010). In recent years, some elaborate methods have been proposed in the literature. Among the first contributions in this context is the work of Li and Ma (2011). To get around interval-censoring, the authors proposed to redefine the ROC but the proposed definition is much less intuitive than the classical one and their estimation method is not efficient as they ignore all observed intervals that include the time horizon t. Using the classical definition, Wu et al. (2020) proposed a non-parametric estimator for the ROC and the AUC based on a nonparametric sieve maximum likelihood estimator of the joint distribution of the survival time and the marker. The authors considered the case where the subjects under study are all censored and assumed joint independence between the observed times

and the pair of the survival time and the marker. Moreover, the authors (without declaring it) assumed that, whatever the time horizon t is, the event status $I(T \leq t)$ is unknown for all subjects under study, which naturally induce loss of information. Closely related papers to this one are the work of Wu and Cook (2020), who considered the evaluation of a parametric prediction model using a validation sample of individuals subject to the case K interval censoring, and the work of Díaz-Coto et al. (2020) who proposed an imputation method based on the semiparametric proportional hazards model to estimate the AUC from a sample of individuals all subject to interval-censoring.

In this chapter, we propose both an empirical (non-smoothed) and a smoothed time-dependent ROC curve estimators for the general case of mixed data that may involve exact, left-censored, right-censored, and interval-censored observations. In this sense, this chapter generalizes the paper of Beyene and El Ghouch (2020b), which presented similar methods for the right-censored data. In our approach, taking the nature of the data into account, only unknown event status are imputed using a weight function that depends on a survival function that needs to be estimated. To this end, we use a nonparametric, semiparametric and parametric approaches. In addition to the AUC, to summarize the ROC curve and to obtain an optimal cutoff value, we studied the Youden index (Youden; 1950). These estimators only require conditional independence between the censoring times and the event time given the marker. As for the variance estimation and inference method, we consider a nonparametric data-based bootstrap approach.

The rest of this chapter is organized as follows. In the next section, we introduce the definitions and propose different estimation methods for the time-dependent sensitivity, specificity, ROC curve, AUC, Youden index, and the corresponding optimal cutoff point from mixed-case interval-censored data with or without completely observed subjects. In this same section, we present a bootstrap approach for estimating the variance and construct confidence intervals. In Section 4.3, we present the results of a simulation study. In Section 4.4, the proposed methodology is illustrated using two real data examples. Finally, we conclude this chapter with a discussion in Section 4.5.

4.2 Method

In this section, we first introduce some important notations and definitions, after that, we will provide estimators for the time-dependent true positive rate, false positive rate, ROC curve, and the corresponding summary measures.

4.2.1 Definitions

Let T denote the time to the event (survival time) of interest and M denote a quantitative marker, with higher value associated with higher risk of getting the event. For a given time horizon t, the event status of a subject is defined as $D(t) = I(T \leq t)$, with 1 indicating the presence of the event before t and 0 otherwise. For a given marker value (cutoff) $m \in \mathbb{R}$, the time-dependent true positive (false positive) rate TPR_t (FPR_t) is the conditional probability that M is greater than m given the survival time T is smaller (greater) than t, thus

$$\begin{aligned} \mathrm{TPR}_t(m) &= P(M > m | D(t) = 1), \\ \mathrm{FPR}_t(m) &= P(M > m | D(t) = 0). \end{aligned} \tag{4.1}$$

The time-dependent ROC curve (ROC_t) can be obtained by plotting TPR_t versus FPR_t for all possibles values of m. Alternatively, the ROC curve may be defined as the graph of the *ROC function* given by

$$\mathrm{ROC}_t(u) = \mathrm{TPR}_t\{\mathrm{FPR}_t^{-1}(u)\}, \tag{4.2}$$

where $u \in [0, 1]$ and $\mathrm{FPR}_t^{-1}(u) = m$ is such that $u = \mathrm{FPR}_t(m)$.

As mentioned before, the two most used summary indices derived from the ROC are the area under the ROC curve (AUC_t), defined as $\mathrm{AUC}_t = \int_0^1 \mathrm{ROC}_t(u)du$, and the Youden index. To define the latter, lets first define the time-dependent *Youden function*, which is nothing but the difference between TPR_t and FPR_t; thus, for a given marker value m, $\mathtt{J}_t(m) = \mathrm{TPR}_t(m) - \mathrm{FPR}_t(m)$. The maximum of this function is the well known Youden index that will be denoted, hereafter, by $\mathtt{J}_t$. Thus,

$$\mathtt{J}_t = \max_m \mathtt{J}_t(m) = \max_{0 \leq u \leq 1} \{\mathrm{ROC}_t(u) - u\}. \tag{4.3}$$

The value of the marker leading to $\mathtt{J}_t$, i.e., $m_t^* = arg\, max_m \mathtt{J}_t(m)$, is the optimal cutoff value in the sense that it optimizes the markers differentiating ability when equal weight is given to sensitivity and specificity.

4.2.2 ROC estimation

In this section, we first derive theoretical formulas for the quantities defined above which serve as the basis for our proposal. As mentioned in the Introduction, we are interested in the mixed case of exact, right, left and interval-censored data. Usually analyzing such data is not an easy task and, typically, researchers focus on one some specific situation like for example the well-studied case of right-censored data. However, as it will be shown below, mixing all these cases become straightforward through our approach.

Let $L \in [0, \infty)$ and $R \in (0, \infty]$ denote two random variables such that $L \leq R$.

These random variables will play the role of, respectively, the left and the right limits of the observed interval time. The survival time T is not observed but only known to be somewhere between L and R. More precisely, we assume that our observations come from the i.i.d. sample $(L_i, R_i, M_i), i = 1, 2..., n$, with the following conventions: the event time T_i is (i) observed if $L_i = R_i$, (ii) left censored if $L_i = 0$ and $R_i < \infty$, (iii) right censored if $L_i > 0$ and $R_i = \infty$ and (iv) interval censored if $0 < L_i < R_i < \infty$. In the following, the vector (L, R) of censoring times is assumed to be independent of T given the marker M.

Let $S(\cdot|M)$ denote the conditional survival probability of $T|M$. Using the assumption above and the "tower property" of conditional expectations, it is easy to see that TPR_t and FPR_t, as given by (4.1), can be written as

$$\mathrm{TPR}_t(m) = \frac{E\{I(M > m)W(t)\}}{E\{W(t)\}}, \ \mathrm{FPR}_t(m) = \frac{E\{I(M > m)(1 - W(t))\}}{E\{1 - W(t)\}}, \quad (4.4)$$

where $W(t) \equiv W_t(L, R, M) = P(T \leq t|L, R, M)$. Here, we have two cases: (1) Uncensored $\Leftrightarrow L = R \Leftrightarrow T = L = R$ and (2) Censored $\Leftrightarrow L < R \Leftrightarrow L < T < R$. In the last case we may distinguish three situations: Right censoring $\Leftrightarrow R = +\infty \Leftrightarrow T > L$, Left censoring $\Leftrightarrow L = 0 \Leftrightarrow T < R$, and Interval censoring $\Leftrightarrow R < +\infty$ and $L > 0 \Leftrightarrow L < T < R$.

For the uncensored case,

$$W(t) = P(T \leq t|T = L = R, Z) = I(T \leq t) = I(R \leq t) = I(L \leq t).$$

For the censored case,

$$\begin{aligned} W(t) &= \frac{P(T \leq t, L < T < R|L, R, M)}{P(L < T < R|L, R, M)} \\ &= \frac{P(T \leq t, L < T < R|M)}{P(L < T < R|M)} \\ &= \frac{P(T \leq t, L < T < R|M)}{P(L < T < R|M)} I(L < t) \\ &= \left\{ \frac{P(L < T \leq t|M)}{P(L < T < R|M)} I(R > t) + I(R \leq t) \right\} I(L < t). \end{aligned}$$

So, combining this two cases, the weight can be expressed as

$$W(t) = I(R \leq t) + \left\{ \frac{S(L|M) - S(t|M)}{S(L|M) - S(R|M)} \right\} I(L < t < R). \quad (4.5)$$

It is interesting to see that this formula reduces to known results for certain special cases. For example, in the case of an exact observation, i.e., $L = T = R$, $W(t)$ becomes simply $I(T \leq t)$ and for a right censored observation, i.e., $L < T < R = \infty$,

$W(t)$ becomes $\left[1 - \frac{S(t|M)}{S(L|M)}\right] I(L < t)$. Which coincides with the formula that can be found in Beyene and El Ghouch (2020b) (see also Li et al. (2018) and Martínez-Camblor et al. (2016)).

It is clear that the formula above can not be used without knowing $S(\cdot|M)$, the conditional survival probability of $T|M$. So we need first to estimate this quantity, which can be done via a multitude of ways some of them are discussed in the next section. For now, let's assume that we are given an estimator $\hat{S}$ for S. By plugging $\hat{S}$ into $W(t)$ we get $\hat{W}(t)$ which can be then used to estimate all our quantities of interest. In fact, a natural estimators for TPR_t and FPR_t, as defined in (4.4), are given by

$$\widehat{TPR}_t(m) = \frac{\sum_{i=1}^n I(M_i > m)\hat{W}_{ti}}{\sum_{i=1}^n \hat{W}_{ti}}, \text{ and } \widehat{FPR}_t(m) = \frac{\sum_{i=1}^n I(M_i > m)(1 - \hat{W}_{ti})}{\sum_{i=1}^n (1 - \hat{W}_{ti})}, \tag{4.6}$$

respectively, where $\hat{W}_{ti} \equiv \hat{W}_i(t) = I(R_i \leq t) + \left\{ \frac{\hat{S}(L_i|M_i) - \hat{S}(t|M_i)}{\hat{S}(L_i|M_i) - \hat{S}(R_i|M_i)} \right\} I(L_i < t < R_i)$. The same approach can be applied to estimate the ROC. In fact, setting $Z(t)$ to be the random variable $FPR_t(M)$, based on (4.2), we have that

$$\text{ROC}_t(u) = E\left\{\mathcal{W}(t) I(Z(t) \leq u)\right\}, \tag{4.7}$$

where $\mathcal{W}(t)$ denotes the ratio $W(t)/E(W(t))$. This naturally suggests the following estimator for the ROC.

$$\widehat{\text{ROC}}_t(u) = n^{-1} \sum_{i=1}^n \hat{\mathcal{W}}_{ti} I(\hat{Z}_{ti} \leq u), \tag{4.8}$$

where $\hat{Z}_{ti} \equiv \hat{Z}_i(t) = \widehat{FPR}_t(M_i)$ and $\hat{\mathcal{W}}_{ti} = \hat{W}_{ti}/\bar{W}_t$, with $\bar{W}_t = n^{-1} \sum_{i=1}^n \hat{W}_{ti}$. Integrating this last formula leads to the following estimate for the AUC.

$$\widehat{\text{AUC}}_t = 1 - n^{-1} \sum_{i=1}^n \hat{\mathcal{W}}_{ti} \hat{Z}_{ti}. \tag{4.9}$$

Finally, the Youden function, at the prediction time t, can be estimated by

$$\hat{\mathtt{J}}_t(m) = \widehat{TPR}_t(m) - \widehat{FPR}_t(m) = n^{-1} \sum_i \frac{1 - \hat{\mathcal{W}}_{ti}}{1 - \bar{W}_t} I(M_i \leq m). \tag{4.10}$$

Maximizing this step function, with respect to m, is a straightforward task and results into an estimate for both the Youden index and the corresponding optimal cutoff value defined above.

The ROC curve estimator given in equation (4.8) generalizes, to mixed interval-censored data, the empirical estimator proposed by Beyene and El Ghouch (2020b) for right-censored data and, so, it inherits the drawback of the latter, namely being a step function. As invoked previously and as it will be demonstrated in our simulations, smoothing can lead to better (more precise) estimates of both the ROC and the optimal cutoff value. To smooth the ROC, Beyene and El Ghouch (2020b) propose the following estimator

$$\widehat{\mathrm{ROC}}_{t,h}(u) = n^{-1}\sum_{i=1}^{n}\hat{\mathcal{W}}_{ti}K\left(\frac{Q(u)-Q(\hat{Z}_{ti})}{h}\right), \tag{4.11}$$

where h is a smoothing parameter, $K(x)=\int_{-\infty}^{x}k(s)ds$ is a (kernel) distribution function with a density k and $Q(\cdot)$ is any given quantile-transformation function (e.g., standard normal quantile function) needed to correct for the well-known kernel-boundary problems caused by the fact that $\hat{Z}_{ti}\in[0,1]$. In our case, $\widehat{\mathrm{ROC}}_{t,h}$ can still be used as it is, but now with the $\hat{\mathcal{W}}_{ti}$'s defined above.

The Youden index, see equation (4.3), can also be estimated by the Kernel method using

$$\mathtt{J}_{t,h} = \max_{0\le u\le 1}\{\widehat{\mathrm{ROC}}_{t,h}(u)-u\}. \tag{4.12}$$

Numerical optimization methods could be used to find the maximum in (4.12). The corresponding optimal cutoff value can be obtained by linear interpolation of $\widehat{\mathrm{TPR}}$ between the points $M_{(\lfloor nu^*_{t,h}\rfloor)}$ and $M_{(\lfloor nu^*_{t,h}\rfloor+1)}$, where $M_{(i)}$ denote the ith order statistic of $M_i, i=1,\ldots,n$, $u^*_{t,h}=arg\,max_{0\le u\le 1}\{\widehat{\mathrm{ROC}}_{t,h}(u)-u\}$, and $\lfloor x\rfloor$ is the floor function.

4.2.3 Conditional survival function estimation

As mentioned above, to apply our method one needs to estimate S, the conditional survival function of $T|M$. In the presence of interval-censored subjects, several approaches are available in the literature, some of them will be briefly reviewed here.

Most used methods in practice are based on (semi-)parametric models like the Cox proportional hazards model and the accelerated failure time (AFT) model. For a given set of covariates $\boldsymbol{X}$, the latter is given by $S_{AFT}(t|\boldsymbol{X}) = S_0(t\exp(\boldsymbol{\beta}'\boldsymbol{X}))$, whereas the former can be expressed as $S_{COX}(t|\boldsymbol{X}) = S_0(t)^{\exp(\boldsymbol{\beta}'\boldsymbol{X})}$. Contrary to the AFT model where the baseline survival function S_0 needs to be parameterized, in the Cox model it is left unspecified. This makes the latter relatively more robust but also more computationally demanding. Note also that there is no intercept in the vector of coefficients $\boldsymbol{\beta}$ in the Cox model. In the literature, to obtain the maximum likelihood estimate of the parameters of these models, several algorithms

(self-consistency algorithm, iterative convex minorant, EM, Newton-Raphson, etc) have been proposed; see Chen et al. (2012) for a book-length review on this subject. Some of these algorithms and new improved variants have been implemented in the `R` package `icenReg` (Anderson-Bergman; 2017b). As with any parametric based approach, the effect of a misspecified model on the resulting ROC (and summary measures) can be huge and may result in a considerable bias. It is of course advocated to check the validity of any statistical model before using it but this is not always possible especially in our context with incomplete data and, typically, small samples sizes.

A more robust alternative could be to use nonparametric techniques. Very few estimators have been proposed and studied in the literature; see Dehghan and Duchesne (2011b) for some examples. Among them, there is the estimator of Dehghan and Duchesne (2011a) who proposed a local kernel version of the well known Turnbull's estimator and suggest an algorithm, that can be seen as an EM-type algorithm, to compute it. We investigated this method via a simulation study (not shown here for the sake of brevity) and find some unsatisfactory results that may be partially explained by the difficulties to select a good bandwidth parameter needed for kernel-smoothing. A completely different approach has be recently proposed by Wu et al. (2020). Following an idea similar to that of Wu and Zhang (2012), these authors developed a tensor spline-based sieve maximum likelihood estimation method to estimate both the joint cumulative distribution function of (T, M) and the marginal distribution of M. Under the independence assumption between (L, R) and (T, M), Wu et al. (2020) have shown the consistency of their estimator and used it to produce nonparametric estimates for the ROC and AUC. The method of Wu et al. (2020) could be easily updated to produce a spline-based estimate for the conditional survival function of $T|M$ which then can be used as a building block for our method as described above. More details will be given in the simulation section where we investigate both this approach and model-based approaches (AFT and COX).

4.2.4 Bootstrap and inference

We can see from the literature, e.g., Heagerty et al. (2000), Wu et al. (2020), Rota et al. (2015), Hall et al. (2004) and Li and Ma (2011), that the Bootstrap (Efron; 1979) is typically advocated to make inference about the ROC and its summary measures. This is because Gaussian approximation, when it is available, can hardly be used as it involves conditional densities that are not easy to estimate nonparametrically; see Chapter 5 of Pepe (2003) for more about this subject. Getting useful asymptotic distribution is even more difficult in our case due to the censoring mechanism. This is why, here, we naturally turn to the Bootstrap.

Hereafter, we will focus on the AUC, but the same procedure can be applied to

each points on the ROC and Youden index estimators. The bootstrap procedure involves drawing a large number, say B, of random samples of size n with replacement from the original data $\{(M_i, L_i, R_i), i = 1, 2, ..., n\}$. From each bootstrap sample b, $b = 1, \ldots, B$, we calculate $\widehat{AUC}_b$, the estimated bootstrap AUC statistic. The empirical bootstrap variance $S_B^2 = B^{-1} \sum_b \left(\widehat{AUC}_b - B^{-1} \sum_b \widehat{AUC}_b\right)^2$, computed from these bootstrap statistics, can be used as an estimate of the variance of $\widehat{AUC}$. This variance can be used to construct the standard bootstrap confidence interval given by $\widehat{AUC} \pm z_{1-\alpha/2} S_B$, where z_α is the $\alpha-$quantile of a $N(0, 1)$. Many other (better) alternative methods of forming bootstrap confidence intervals exist in the literature; see, for example, Horowitz (2001). Among them there is the percentile method given by $[\widehat{AUC}_B(\alpha/2), \widehat{AUC}_B(1 - \alpha/2)]$, where $\widehat{AUC}_B(\alpha)$ is the $(B\alpha)^{\text{th}}$ order statistic of the bootstrap estimates $\{\widehat{AUC}_b,\ b = 1, \ldots, B\}$. In the next section, these two methods will be compared.

4.3 Simulation study

In this section, we conduct a Monte Carlo simulation study to examine the finite sample properties of the proposed time-dependent ROC curve and the associated summary measures. Another objective of this study is to compare the smooth and non-smooth versions of our estimators and also to compare our method with the one proposed by Wu et al. (2020). The next section present the data generating process, the results will follow after.

4.3.1 Data generation

We generate 1000 data of size $n = 100, 200$ according to the the following procedure:

1. We generate the survival time T_i from a Weibull distribution with shape and scale parameters given, respectively, by $\kappa = 1.5$ and $\lambda = 2.35$.

2. Let $\xi_i \sim Unif(0.75, 1)$ and $\nu_i \sim Unif(1, r)$, where $r > 1$. Define the left and right censoring times to be $L_i = \xi_i T_i$ and $R = \nu_i T_i$, respectively. r controls the length of the observed intervals and will be chosen to achieve an average length (AL) of either 0.75, 1.5, or 3.

3. We generate the marker according to $M_i = \sqrt{\rho} T_i + \sqrt{1 - \rho}\epsilon_i$, where ϵ_i is a Weibull distributed random variable with parameters $\kappa = 1.5$ and $\lambda = 2.35$. The values of ρ are set to be 0.25 and 0.5. The latter corresponds to a stronger association between the survival time and the marker.

All observations generated in this way are interval-censored. To incorporate right and/or left censoring we consider the following two cases:

- **Case I (mixed case of right and interval censored subjects)**: Let $\Delta_i^R \sim Bernoulli(p_R)$ with $p_R = 0.1, 0.2, 0.3$. If, for a given i, $\Delta_i^R = 1$ then R_i is set to ∞ and the observation becomes right censored; otherwise, the observation i remains interval-censored.

- **Case II (mixed case of left, right and interval censored subjects)**: Let $(\Delta_i^R, \Delta_i^L, \Delta_i^I) \sim Multinomial(1, (0.1, p_L, 1-0.1-p_L))$ with $p_L = 0.1, 0.2$, 0.3. If, for a given i, (i) $\Delta_i^R = 1$ then R_i is set to ∞ and the observation becomes right censored, (ii) $\Delta_i^L = 1$ then L_i is set to 0 and the observation becomes left censored; otherwise, the observation i remains interval-censored.

As time horizon t, we consider the quartile values (Q_1, Q_2 and Q_3) of the survival time T. The true ROC curves and the corresponding AUC and Youden function are presented in Figure 4.1. The true ROC was calculated from a simulated sample (T_i, M_i), $i = 1, ..., 4 \times 10^6$, using equation (4.8) with $\hat{W}_i = I(T_i \leq t)$.

The tables provided in the next sections show, among other information, the integrated mean bias (MIB) and the integrated mean square error (MISE) given, respectively, by

$$\text{MIB}(\widehat{\text{ROC}}) = 1000^{-1} \sum_{l=1}^{1000} \int (\widehat{\text{ROC}}_l(u) - \text{ROC}(u))du,$$

and

$$\text{MISE}(\widehat{\text{ROC}}) = 1000^{-1} \sum_{l=1}^{1000} \text{ISE}(\widehat{\text{ROC}}_l),$$

where $\widehat{\text{ROC}}_l(.)$ is the ROC estimate from the l^{th} sampled data and $\text{ISE}(\widehat{\text{ROC}}_l) = \int (\widehat{\text{ROC}}_l(u) - \text{ROC}(u))^2 du$.

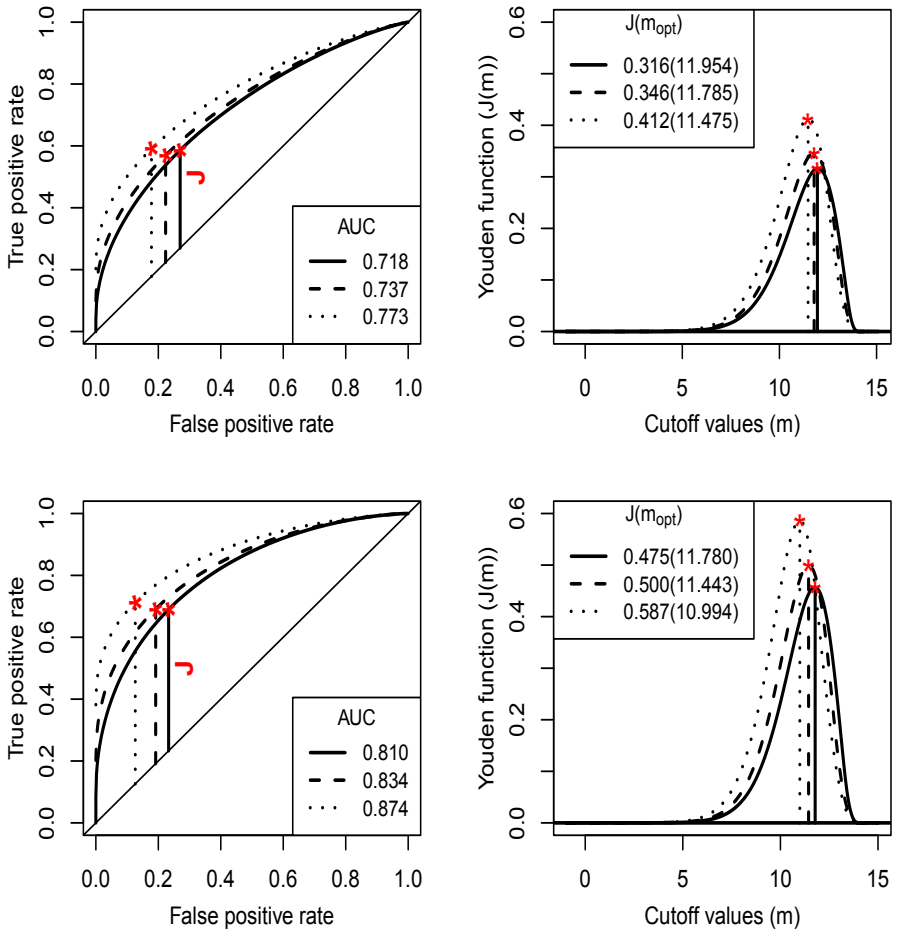

Figure 4.1: The true ROC curve and the corresponding AUC and Youden function for $\rho = 0.25$ (top row) and for $\rho = 0.5$ (bottom row) for the prediction time Q_1 (solid line), Q_2 (dashed line) and Q_3 (dotted line).

4.3.2 Simulation results for the ROC

In this section, we present and discuss the main findings of our simulation study. We will start by comparing our method with the one of Wu et al. (2020) (hereafter referred by WU) as implemented in the `R` package `intcensROC` (Lin et al.; 2018). This package also provides a spline estimate for the joint distribution of (T, M) and for the marginal distribution of M. We use this and numerical differentiation

to get an estimate for the conditional distribution $T|M$ needed in the calculation of W_i's; see equation (4.5). In our calculation, we use cubic splines and follow the recommendations of Wu et al. (2020) regarding the number of the knot and their placements. To estimate the conditional distribution $T|M$, in addition to the spline method, we also use the parametric AFT and semiparametric COX models. The performances of the resulting ROC estimates (AFT, COX, Spline, WU) are compared in terms of bias and mean squared error. For the AFT approach we considered Gamma, Log-normal and Weibull distribution as baseline survival function.

Table 4.1 shows the obtained results for **Case I** with $\rho = 0.5$ (results for $\rho = 0.25$ are very similar and therefor are not shown here). Let's first focus on the empirical ROC, see equation (4.8), based on the AFT model. When comparing the performances of the different baseline survival functions (Gamma, Log-normal, and Weibull), we can observe somewhat similar results with a small advantage to the Weibull. Globally, the resulting ROC performs slightly better than the one based on the semi-parametric Cox model. Both the AFT and Cox based ROC estimators tend to do better than the corresponding spline based ROC. The latter, however, outperforms the WU estimator who has both larger MIB and larger MISE. In general, this applies to all censoring rates, interval lengths and sample sizes. Very similar comments apply to Table 4.2 that show some of the obtained results for **Case II**. Moreover, both in these two tables, it can be seen that as the sample size increases both the MIB and the MISE decrease which demonstrates consistency. It is also worth noting that increasing the rate of (left or right) censoring or increasing the length of the observed intervals affects badly all the estimates.

In Table 4.3, we compare the empirical (non-smooth) ROC estimator (4.8) and the smooth ROC estimator (4.11). Again we don't show all the obtained results but only some representative cases. To choose the bandwidth parameter h we use the adapted normal reference (NR) method proposed by Beyene and El Ghouch (2020b). Among all the methods proposed by theses authors, we choose the NR method for its simplicity and because it has been shown to produce, relatively, better results. From Table 4.3, one can clearly see that the proposed smooth ROC estimator has systematically smaller MISE than the non-smooth version. The latter seems, however, to perform better in terms of bias. This is, generally, true for all the considered scenarios. This is consistent with the finding of a simulation study by Beyene and El Ghouch (2020b) conducted in the context of right-censored data.

Table 4.1: The MIB($\times 10^{-2}$) and MISE($\times 10^{-2}$) of the proposed empirical ROC estimator and the WU method computed for different sample sizes (n), average intervals lengths (AL), right censoring rates (%RC), and t values. **Case I,** $\rho = 0.5$.

			AFT											
			Gamma		Lognormal		Weibull		COX		Spline		WU	
AL	%RC	t	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE
$n = 100$														
0.75	10	1.0	0.218	0.587	0.107	0.582	0.361	0.593	0.032	0.629	0.305	0.633	2.068	0.630
0.75	10	1.8	0.437	0.358	0.713	0.370	0.224	0.354	-0.208	0.377	0.876	0.433	8.594	1.440
0.75	10	2.9	1.165	0.288	2.067	0.394	0.403	0.243	-0.454	0.269	3.494	0.820	19.40	6.269
1.50	10	1.0	0.244	0.632	-0.047	0.622	0.548	0.647	0.045	0.720	0.348	0.718	2.099	0.689
1.50	10	1.8	0.775	0.369	1.030	0.384	0.539	0.362	-0.254	0.394	0.586	0.452	6.863	1.160
1.50	10	2.9	1.959	0.358	2.965	0.497	0.850	0.266	-0.654	0.276	4.703	0.886	22.38	7.374
3.00	10	1.0	0.158	0.739	-0.455	0.720	0.727	0.774	-0.255	0.908	0.453	0.892	2.001	0.778
3.00	10	1.8	0.926	0.410	1.105	0.425	0.603	0.397	-0.881	0.451	-0.198	0.501	3.903	0.871
3.00	10	2.9	2.552	0.477	3.723	0.668	0.923	0.310	-1.330	0.331	5.640	1.007	20.43	5.951
0.75	20	1.0	0.357	0.626	0.232	0.618	0.532	0.636	0.069	0.678	0.667	0.706	2.930	0.767
0.75	20	1.8	0.729	0.385	1.188	0.413	0.364	0.371	-0.178	0.390	1.925	0.574	9.719	1.842
0.75	20	2.9	1.760	0.359	3.230	0.590	0.579	0.258	-0.437	0.269	7.929	2.061	25.64	9.585
1.50	20	1.0	0.347	0.678	0.031	0.662	0.684	0.697	0.001	0.791	0.671	0.797	2.852	0.817
1.50	20	1.8	0.936	0.399	1.303	0.424	0.585	0.383	-0.379	0.411	1.238	0.554	7.652	1.444
1.50	20	2.9	2.334	0.427	3.695	0.652	0.921	0.286	-0.803	0.287	8.110	1.900	25.78	9.314
3.00	20	1.0	0.228	0.814	-0.422	0.778	0.821	0.847	-0.318	1.019	0.685	0.987	2.626	0.908
3.00	20	1.8	0.983	0.442	1.214	0.462	0.580	0.422	-1.043	0.483	0.256	0.584	4.777	1.107
3.00	20	2.9	2.757	0.536	4.103	0.775	0.919	0.330	-1.388	0.350	7.357	1.562	21.48	6.664
0.75	30	1.0	0.564	0.679	0.420	0.668	0.771	0.693	0.150	0.733	1.175	0.810	3.842	0.928
0.75	30	1.8	1.105	0.425	1.745	0.480	0.586	0.395	-0.109	0.410	3.285	0.803	10.53	2.183
0.75	30	2.9	2.395	0.457	4.397	0.845	0.782	0.280	-0.460	0.275	12.14	3.589	28.28	11.05
1.50	30	1.0	0.441	0.744	0.098	0.726	0.822	0.771	-0.053	0.897	1.063	0.918	3.614	0.969
1.50	30	1.8	1.157	0.440	1.643	0.485	0.705	0.414	-0.458	0.445	2.290	0.730	8.387	1.737
1.50	30	2.9	2.707	0.505	4.432	0.839	0.995	0.308	-0.793	0.301	11.31	3.069	26.73	9.865
3.00	30	1.0	0.204	0.892	-0.463	0.862	0.850	0.947	-0.504	1.136	0.961	1.128	3.310	1.056
3.00	30	1.8	1.077	0.485	1.358	0.514	0.601	0.459	-1.224	0.532	1.125	0.715	5.718	1.380
3.00	30	2.9	2.965	0.608	4.488	0.905	0.903	0.357	-1.539	0.380	9.092	2.220	21.27	6.730
$n = 200$														
0.75	10	1.0	0.151	0.311	0.057	0.309	0.287	0.315	0.020	0.322	0.391	0.335	2.729	0.432
0.75	10	1.8	0.223	0.179	0.525	0.191	0.002	0.175	-0.271	0.186	-0.171	0.209	3.010	0.496
0.75	10	2.9	0.790	0.158	1.717	0.253	0.035	0.126	-0.637	0.142	2.941	0.524	18.52	5.547
1.50	10	1.0	0.156	0.332	-0.118	0.326	0.443	0.341	-0.058	0.349	0.483	0.375	2.535	0.450
1.50	10	1.8	0.612	0.201	0.887	0.217	0.358	0.191	-0.220	0.202	-0.267	0.253	2.373	0.536
1.50	10	2.9	1.638	0.225	2.670	0.356	0.513	0.143	-0.659	0.148	3.446	0.468	17.08	4.401
3.00	10	1.0	-0.008	0.379	-0.596	0.369	0.512	0.398	-0.387	0.414	0.336	0.464	2.038	0.491
3.00	10	1.8	0.801	0.243	0.998	0.259	0.431	0.223	-0.687	0.238	0.184	0.352	2.488	0.672
3.00	10	2.9	2.382	0.354	3.576	0.540	0.695	0.185	-1.136	0.183	1.906	0.314	10.31	2.029
0.75	20	1.0	0.338	0.326	0.219	0.322	0.505	0.332	0.099	0.339	0.696	0.365	3.417	0.522
0.75	20	1.8	0.560	0.200	1.045	0.229	0.173	0.185	-0.197	0.193	1.065	0.310	5.151	0.909
0.75	20	2.9	1.419	0.222	2.941	0.442	0.210	0.134	-0.610	0.142	7.052	1.578	22.95	7.748
1.50	20	1.0	0.269	0.352	-0.032	0.343	0.599	0.365	-0.072	0.368	0.660	0.410	3.095	0.528
1.50	20	1.8	0.798	0.223	1.202	0.252	0.439	0.205	-0.290	0.213	0.759	0.340	4.395	0.875
1.50	20	2.9	2.031	0.286	3.468	0.512	0.603	0.157	-0.702	0.154	5.990	1.157	19.10	5.585
3.00	20	1.0	0.032	0.403	-0.580	0.391	0.598	0.427	-0.488	0.446	0.389	0.503	2.444	0.554
3.00	20	1.8	0.871	0.264	1.147	0.288	0.441	0.239	-0.809	0.253	0.910	0.428	4.169	0.936
3.00	20	2.9	2.596	0.407	4.022	0.655	0.710	0.201	-1.267	0.192	3.315	0.727	12.11	3.046
0.75	30	1.0	0.635	0.365	0.503	0.358	0.838	0.375	0.266	0.372	1.192	0.437	4.104	0.656
0.75	30	1.8	0.933	0.237	1.614	0.292	0.384	0.206	-0.118	0.205	2.683	0.527	7.210	1.434
0.75	30	2.9	2.048	0.313	4.120	0.685	0.399	0.152	-0.596	0.148	10.70	2.856	23.64	8.195
1.50	30	1.0	0.492	0.393	0.159	0.380	0.861	0.411	-0.004	0.410	1.013	0.483	3.630	0.649
1.50	30	1.8	1.056	0.260	1.568	0.303	0.563	0.227	-0.341	0.228	2.197	0.515	6.418	1.326
1.50	30	2.9	2.477	0.374	4.245	0.696	0.691	0.177	-0.761	0.167	8.443	2.059	19.12	5.945
3.00	30	1.0	0.168	0.451	-0.475	0.433	0.783	0.484	-0.514	0.505	0.567	0.585	2.829	0.668
3.00	30	1.8	0.996	0.296	1.337	0.331	0.492	0.263	-0.972	0.279	2.024	0.562	5.966	1.292
3.00	30	2.9	2.841	0.473	4.469	0.787	0.730	0.222	-1.327	0.210	5.491	1.467	13.65	4.072

Table 4.2: The MIB($\times 10^{-2}$) and MISE($\times 10^{-2}$) of the proposed empirical ROC estimators and the WU's method computed for different left censoring rates (%LC), sample sizes (n), ρ and t values. **Case II,** $AL = 0.75$.

			AFT											
			Gamma		Lognormal		Weibull		COX		Spline		WU	
ρ	%LC	t	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE
$n = 100$														
0.25	10	1.0	1.251	0.688	1.040	0.689	1.420	0.690	1.173	0.722	2.064	0.812	6.916	1.101
0.25	10	1.8	1.258	0.477	1.503	0.488	1.177	0.474	0.805	0.495	2.245	0.560	8.843	1.308
0.25	10	2.9	1.672	0.467	2.591	0.536	1.143	0.446	0.354	0.468	3.508	0.727	13.51	3.138
0.25	20	1.0	2.450	0.677	2.151	0.683	2.632	0.679	2.312	0.693	4.843	0.997	10.41	1.732
0.25	20	1.8	2.029	0.514	2.345	0.542	1.909	0.503	1.345	0.515	3.906	0.686	11.00	1.802
0.25	20	2.9	2.187	0.510	3.226	0.606	1.636	0.479	0.721	0.498	4.804	0.909	14.64	3.546
0.25	30	1.0	3.671	0.686	3.267	0.692	3.807	0.682	3.309	0.670	7.628	1.321	12.86	2.283
0.25	30	1.8	3.159	0.586	3.532	0.635	2.995	0.567	2.247	0.551	6.073	0.943	12.99	2.355
0.25	30	2.9	2.904	0.592	4.056	0.720	2.356	0.551	1.205	0.559	6.276	1.190	15.75	3.963
0.50	10	1.0	1.340	0.557	1.163	0.555	1.537	0.563	0.971	0.581	2.776	0.706	10.04	1.699
0.50	10	1.8	1.391	0.371	1.734	0.398	1.136	0.357	0.385	0.367	2.879	0.522	12.64	2.421
0.50	10	2.9	1.745	0.322	2.726	0.454	0.973	0.264	-0.226	0.273	5.001	1.053	20.60	6.906
0.50	20	1.0	2.660	0.600	2.375	0.598	2.876	0.605	2.069	0.588	6.642	1.156	15.01	3.126
0.50	20	1.8	2.326	0.435	2.717	0.483	2.020	0.406	0.901	0.392	5.324	0.797	15.82	3.522
0.50	20	2.9	2.350	0.375	3.436	0.547	1.547	0.298	0.037	0.286	6.787	1.439	21.99	7.661
0.50	30	1.0	4.006	0.664	3.562	0.656	4.199	0.662	2.954	0.584	10.55	1.897	18.58	4.435
0.50	30	1.8	3.528	0.533	3.940	0.605	3.179	0.487	1.613	0.415	8.216	1.284	18.71	4.714
0.50	30	2.9	3.050	0.455	4.244	0.673	2.231	0.359	0.299	0.316	8.669	1.943	23.32	8.435
$n = 200$														
0.25	10	1.0	1.211	0.367	1.027	0.367	1.363	0.370	1.193	0.376	2.432	0.461	7.560	0.952
0.25	10	1.8	0.981	0.245	1.259	0.256	0.884	0.241	0.604	0.251	0.818	0.282	4.976	0.569
0.25	10	2.9	1.142	0.242	2.108	0.304	0.594	0.223	-0.030	0.234	3.078	0.416	13.74	2.941
0.25	20	1.0	2.253	0.362	1.979	0.362	2.409	0.366	2.132	0.364	5.034	0.656	10.65	1.576
0.25	20	1.8	1.722	0.264	2.066	0.287	1.584	0.255	1.172	0.256	2.221	0.327	7.396	0.899
0.25	20	2.9	1.605	0.265	2.672	0.347	1.053	0.239	0.345	0.246	4.153	0.534	14.80	3.320
0.25	30	1.0	3.291	0.414	2.923	0.410	3.407	0.415	3.005	0.393	7.419	0.997	12.64	2.098
0.25	30	1.8	2.581	0.322	2.987	0.363	2.404	0.307	1.848	0.287	3.986	0.475	9.650	1.356
0.25	30	2.9	2.250	0.323	3.440	0.438	1.703	0.288	0.732	0.279	5.559	0.759	16.05	3.824
0.50	10	1.0	1.289	0.306	1.136	0.304	1.465	0.311	0.999	0.304	3.155	0.448	10.69	1.650
0.50	10	1.8	1.143	0.196	1.511	0.221	0.865	0.182	0.270	0.186	1.146	0.229	7.245	0.961
0.50	10	2.9	1.358	0.186	2.370	0.306	0.572	0.137	-0.432	0.144	4.456	0.734	20.07	6.337
0.50	20	1.0	2.490	0.336	2.241	0.333	2.668	0.341	1.957	0.311	6.840	0.881	15.27	3.021
0.50	20	1.8	2.057	0.235	2.472	0.277	1.732	0.209	0.813	0.190	3.049	0.327	10.62	1.650
0.50	20	2.9	1.922	0.219	3.019	0.369	1.124	0.156	-0.117	0.146	5.973	0.996	21.47	7.079
0.50	30	1.0	3.731	0.431	3.347	0.421	3.880	0.432	2.839	0.357	10.38	1.614	18.28	4.191
0.50	30	1.8	3.108	0.321	3.560	0.386	2.738	0.280	1.474	0.219	5.527	0.605	13.74	2.560
0.50	30	2.9	2.648	0.282	3.861	0.478	1.835	0.200	0.200	0.159	7.863	1.435	23.16	8.030

Table 4.3: The MIB($\times 10^{-2}$) and MISE($\times 10^{-2}$) of the proposed empirical and smooth ROC estimators for different sample sizes (n), average intervals lengths (AL), right censoring rates (%RC), and t values. **Case I,** $\rho = 0.5$.

			AFT-Weibull				**COX**				**Spline**			
			Empirical		**Smooth**		**Empirical**		**Smooth**		**Empirical**		**Smooth**	
AL	%RC	t	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**
$n = 100$														
0.75	10	1.0	0.361	0.593	1.656	0.435	0.032	0.629	1.358	0.454	0.305	0.633	1.588	0.464
0.75	10	1.8	0.224	0.354	1.107	0.285	-0.208	0.377	0.688	0.298	0.876	0.433	1.725	0.359
0.75	10	2.9	0.403	0.243	1.023	0.230	-0.454	0.269	0.119	0.245	3.494	0.820	4.048	0.774
1.50	10	1.0	0.548	0.647	1.904	0.474	0.045	0.720	1.440	0.514	0.348	0.718	1.681	0.528
1.50	10	1.8	0.539	0.362	1.450	0.302	-0.254	0.394	0.683	0.316	0.586	0.452	1.457	0.375
1.50	10	2.9	0.850	0.266	1.504	0.269	-0.654	0.276	-0.058	0.264	4.703	0.886	5.328	0.877
3.00	10	1.0	0.727	0.774	2.146	0.558	-0.255	0.908	1.198	0.622	0.453	0.892	1.823	0.650
3.00	10	1.8	0.603	0.397	1.543	0.337	-0.881	0.451	0.096	0.360	-0.198	0.501	0.674	0.404
3.00	10	2.9	0.923	0.310	1.594	0.321	-1.330	0.331	-0.733	0.321	5.640	1.007	6.311	1.027
0.75	20	1.0	0.532	0.636	1.846	0.467	0.069	0.678	1.419	0.489	0.667	0.706	1.963	0.525
0.75	20	1.8	0.364	0.371	1.251	0.306	-0.178	0.390	0.727	0.313	1.925	0.574	2.750	0.498
0.75	20	2.9	0.579	0.258	1.208	0.253	-0.437	0.269	0.146	0.252	7.929	2.061	8.468	2.001
1.50	20	1.0	0.684	0.697	2.058	0.513	0.001	0.791	1.401	0.558	0.671	0.797	2.018	0.595
1.50	20	1.8	0.585	0.383	1.499	0.323	-0.379	0.411	0.570	0.332	1.238	0.554	2.088	0.470
1.50	20	2.9	0.921	0.286	1.580	0.293	-0.803	0.287	-0.206	0.277	8.110	1.900	8.707	1.876
3.00	20	1.0	0.821	0.847	2.266	0.612	-0.318	1.019	1.092	0.681	0.685	0.987	2.072	0.727
3.00	20	1.8	0.580	0.422	1.525	0.360	-1.043	0.483	-0.056	0.387	0.256	0.584	1.107	0.477
3.00	20	2.9	0.919	0.330	1.594	0.344	-1.388	0.350	-0.783	0.340	7.357	1.562	8.000	1.565
$n = 200$														
0.75	10	1.0	0.287	0.315	1.204	0.242	0.020	0.322	0.951	0.244	0.391	0.335	1.303	0.261
0.75	10	1.8	0.002	0.175	0.592	0.145	-0.271	0.186	0.329	0.152	-0.171	0.209	0.404	0.169
0.75	10	2.9	0.035	0.126	0.435	0.120	-0.637	0.142	-0.257	0.131	2.941	0.524	3.324	0.507
1.50	10	1.0	0.443	0.341	1.413	0.263	-0.058	0.349	0.936	0.261	0.483	0.375	1.442	0.293
1.50	10	1.8	0.358	0.191	0.973	0.164	-0.220	0.202	0.409	0.168	-0.267	0.253	0.329	0.207
1.50	10	2.9	0.513	0.143	0.939	0.146	-0.659	0.148	-0.257	0.142	3.446	0.468	3.870	0.472
3.00	10	1.0	0.512	0.398	1.570	0.303	-0.387	0.414	0.716	0.300	0.336	0.464	1.372	0.359
3.00	10	1.8	0.431	0.223	1.072	0.193	-0.687	0.238	-0.017	0.195	0.184	0.352	0.797	0.301
3.00	10	2.9	0.695	0.185	1.140	0.191	-1.136	0.183	-0.720	0.176	1.906	0.314	2.321	0.313
0.75	20	1.0	0.505	0.332	1.445	0.257	0.099	0.339	1.056	0.257	0.696	0.365	1.629	0.287
0.75	20	1.8	0.173	0.185	0.771	0.156	-0.197	0.193	0.412	0.160	1.065	0.310	1.634	0.269
0.75	20	2.9	0.210	0.134	0.619	0.132	-0.610	0.142	-0.221	0.134	7.052	1.578	7.426	1.556
1.50	20	1.0	0.599	0.365	1.593	0.282	-0.072	0.368	0.953	0.273	0.660	0.410	1.642	0.321
1.50	20	1.8	0.439	0.205	1.060	0.177	-0.290	0.213	0.349	0.178	0.759	0.340	1.348	0.293
1.50	20	2.9	0.603	0.157	1.034	0.161	-0.702	0.154	-0.294	0.150	5.990	1.157	6.395	1.152
3.00	20	1.0	0.598	0.427	1.691	0.325	-0.488	0.446	0.657	0.320	0.389	0.503	1.454	0.387
3.00	20	1.8	0.441	0.239	1.087	0.207	-0.809	0.253	-0.128	0.207	0.910	0.428	1.519	0.376
3.00	20	2.9	0.710	0.201	1.159	0.207	-1.267	0.192	-0.847	0.185	3.315	0.727	3.697	0.710

4.3.3 Summary measures estimators and the bootstrap

In this section, we assess the finite sample performances of the proposed time-dependent AUC, see equation (4.9), the proposed empirical (non-smoothed) and smoothed Youden index, and the corresponding optimal cut-points as described at the end of Section 4.2.2. The results are summarized in Table 4.4 for the AUC and in Table 4.5 for the Youden index and the optimal cut-point. Globally, these results confirm what we have learned from the previous simulations about ROC. Thus, the parametric based version of the AUC estimate performs better followed, closely, by our spline-based estimator. Compared to WU method, the proposed estimators improve significantly the fit, both in terms of bias and mean square error. Furthermore, contrary to the WU method, we can see that the performances of our method get, in general, better with the strength of correlation between event-time and marker. In most cases, the overall performance of both the proposed estimators and the WU method decreases as the length of the observed intervals increase. The gain due to smoothing is also very noticeable. In fact, from Table 4.5, we can see that it reduces significantly both the bias and the mean square error of the resulting Youden index and optimal cut-point estimates. Finally, as expected, the results get better (worse) as the sample size (censoring rate) increases.

We also investigated the validity and the performance of the bootstrap method, as described in Section 4.2.4, to construct confidence intervals for the time-dependent AUC. To this end, we drew $B = 1000$ bootstrap samples for each of the simulated data and then calculated the coverage probabilities (CP) and the average widths (AW). The former refers to the proportion of times that the confidence interval encloses the true AUC value and the latter is the average difference between the lower and upper limits of the confidence intervals. Table 4.6 shows some results. From this table, one can clearly see that, regarding the coverage probability, both methods result in values that are very close to the nominal level of 0.95. It can also be seen that the CP improves with sample size. As for precision, it is observed that the average width decreases as the sample size increases in all the cases. Globally, the percentile method seems to be better as it tends to give more narrow confidence intervals.

Table 4.4: Bias ($\times 10^{-2}$) and MSE ($\times 10^{-2}$) of the proposed time-dependent AUC estimators and the WU method computed for different right censoring rates (%RC), average intervals lengths (AL), sample sizes (n), ρ and t values. **Case I**.

			$\rho = 0.25$				$\rho = 0.50$			
			AFT-Weibull	**COX**	**Spline**	**WU**	**AFT-Weibull**	**COX**	**Spline**	**WU**
AL	%RC	t	**Bias(MSE)**	**Bias(MSE)**	**Bias(MSE)**	**Bias(MSE)**	**Bias(MSE)**	**Bias(MSE)**	**Bias(MSE)**	**Bias(MSE)**
$\boldsymbol{n = 100}$										
0.75	10	1.0	0.29(0.37)	0.13(0.40)	0.15(0.40)	1.35(0.41)	0.36(0.25)	0.03(0.27)	0.30(0.27)	2.07(0.31)
0.75	10	1.8	0.38(0.25)	0.11(0.26)	0.79(0.27)	6.14(0.60)	0.22(0.16)	-0.21(0.17)	0.88(0.19)	8.59(0.95)
0.75	10	2.9	0.66(0.21)	0.08(0.23)	2.44(0.28)	12.7(1.94)	0.40(0.10)	-0.45(0.12)	3.49(0.26)	19.4(4.11)
1.50	10	1.0	0.42(0.40)	0.22(0.45)	0.16(0.47)	1.40(0.47)	0.55(0.27)	0.04(0.31)	0.35(0.31)	2.10(0.35)
1.50	10	1.8	0.75(0.25)	0.29(0.27)	0.64(0.29)	4.99(0.51)	0.54(0.16)	-0.25(0.18)	0.59(0.19)	6.86(0.70)
1.50	10	2.9	1.27(0.22)	0.18(0.24)	3.77(0.33)	15.8(2.75)	0.85(0.11)	-0.65(0.13)	4.70(0.35)	22.4(5.31)
3.00	10	1.0	0.62(0.48)	0.19(0.57)	0.31(0.61)	1.41(0.57)	0.73(0.33)	-0.26(0.40)	0.45(0.40)	2.00(0.40)
3.00	10	1.8	0.99(0.27)	0.10(0.30)	-0.36(0.33)	2.48(0.37)	0.60(0.18)	-0.88(0.22)	-0.20(0.19)	3.90(0.36)
3.00	10	2.9	1.60(0.24)	-0.03(0.28)	5.29(0.49)	15.9(2.75)	0.92(0.13)	-1.33(0.17)	5.64(0.48)	20.4(4.43)
0.75	20	1.0	0.44(0.40)	0.24(0.43)	0.36(0.44)	1.90(0.46)	0.53(0.27)	0.07(0.29)	0.67(0.30)	2.93(0.39)
0.75	20	1.8	0.63(0.27)	0.31(0.28)	1.50(0.31)	6.88(0.74)	0.36(0.17)	-0.18(0.18)	1.92(0.24)	9.72(1.19)
0.75	20	2.9	0.99(0.23)	0.33(0.25)	5.78(0.58)	17.8(3.42)	0.58(0.11)	-0.44(0.12)	7.93(0.84)	25.7(6.90)
1.50	20	1.0	0.55(0.43)	0.26(0.51)	0.36(0.51)	1.90(0.52)	0.68(0.29)	0.00(0.34)	0.67(0.34)	2.85(0.42)
1.50	20	1.8	0.87(0.27)	0.31(0.29)	0.95(0.32)	5.35(0.58)	0.59(0.17)	-0.38(0.19)	1.24(0.21)	7.65(0.83)
1.50	20	2.9	1.45(0.24)	0.24(0.26)	6.40(0.63)	18.6(3.70)	0.92(0.12)	-0.80(0.14)	8.11(0.84)	25.8(6.94)
3.00	20	1.0	0.72(0.52)	0.21(0.65)	0.47(0.68)	1.80(0.64)	0.82(0.36)	-0.32(0.45)	0.69(0.44)	2.63(0.48)
3.00	20	1.8	1.01(0.29)	-0.02(0.33)	-0.28(0.35)	2.79(0.41)	0.58(0.19)	-1.04(0.23)	0.26(0.20)	4.78(0.44)
3.00	20	2.9	1.67(0.26)	-0.13(0.31)	6.46(0.66)	16.4(2.90)	0.92(0.14)	-1.39(0.18)	7.36(0.74)	21.5(4.88)
$\boldsymbol{n = 200}$										
0.75	10	1.0	0.24(0.20)	0.11(0.20)	0.23(0.21)	1.90(0.22)	0.29(0.13)	0.02(0.14)	0.39(0.14)	2.73(0.21)
0.75	10	1.8	0.12(0.12)	-0.02(0.13)	-0.30(0.14)	1.91(0.18)	0.00(0.08)	-0.27(0.08)	-0.17(0.08)	3.01(0.18)
0.75	10	2.9	0.09(0.10)	-0.32(0.12)	1.92(0.14)	12.5(1.71)	0.04(0.05)	-0.64(0.06)	2.94(0.15)	18.5(3.58)
1.50	10	1.0	0.35(0.21)	0.10(0.22)	0.32(0.24)	1.82(0.25)	0.44(0.14)	-0.06(0.15)	0.48(0.16)	2.53(0.22)
1.50	10	1.8	0.53(0.13)	0.23(0.14)	-0.45(0.16)	1.37(0.18)	0.36(0.08)	-0.22(0.09)	-0.27(0.09)	2.37(0.16)
1.50	10	2.9	0.76(0.11)	0.03(0.12)	2.46(0.17)	11.9(1.55)	0.51(0.06)	-0.66(0.07)	3.45(0.18)	17.1(3.05)
3.00	10	1.0	0.43(0.24)	-0.03(0.26)	0.26(0.31)	1.47(0.31)	0.51(0.17)	-0.39(0.18)	0.34(0.20)	2.04(0.23)
3.00	10	1.8	0.79(0.14)	0.18(0.15)	0.11(0.19)	1.59(0.22)	0.43(0.09)	-0.69(0.11)	0.18(0.13)	2.49(0.20)
3.00	10	2.9	1.25(0.13)	0.02(0.14)	0.20(0.12)	5.90(0.48)	0.69(0.06)	-1.14(0.09)	1.91(0.10)	10.3(1.17)
0.75	20	1.0	0.41(0.21)	0.24(0.22)	0.39(0.22)	2.40(0.26)	0.50(0.14)	0.10(0.15)	0.70(0.16)	3.42(0.26)
0.75	20	1.8	0.41(0.13)	0.23(0.14)	0.45(0.15)	3.28(0.26)	0.17(0.08)	-0.2(0.09)	1.06(0.10)	5.15(0.37)
0.75	20	2.9	0.48(0.11)	-0.03(0.12)	5.03(0.38)	16.2(2.76)	0.21(0.05)	-0.61(0.06)	7.05(0.60)	22.9(5.42)
1.50	20	1.0	0.48(0.23)	0.15(0.24)	0.39(0.26)	2.24(0.28)	0.60(0.15)	-0.07(0.16)	0.66(0.18)	3.09(0.26)
1.50	20	1.8	0.69(0.14)	0.31(0.15)	0.22(0.16)	2.77(0.24)	0.44(0.08)	-0.29(0.10)	0.76(0.11)	4.40(0.31)
1.50	20	2.9	1.01(0.12)	0.22(0.13)	4.11(0.30)	13.3(1.93)	0.60(0.06)	-0.70(0.07)	5.99(0.45)	19.1(3.79)
3.00	20	1.0	0.49(0.26)	-0.06(0.29)	0.26(0.34)	1.84(0.35)	0.60(0.18)	-0.49(0.19)	0.39(0.22)	2.44(0.27)
3.00	20	1.8	0.84(0.15)	0.16(0.16)	0.63(0.20)	2.87(0.28)	0.44(0.09)	-0.81(0.11)	0.91(0.14)	4.17(0.32)
3.00	20	2.9	1.38(0.13)	0.07(0.15)	0.95(0.13)	7.09(0.64)	0.71(0.07)	-1.27(0.09)	3.32(0.18)	12.1(1.57)

Table 4.5: Bias ($\times 10^{-2}$) and MSE ($\times 10^{-2}$) for the empirical (non-smoothed) and smoothed Youden index estimators computed for different right censoring rates (%RC), average intervals lengths (AL), sample sizes (n), ρ and t values and for the corresponding optimal cut-point bias ($\times 10^{-1}$) and MSE ($\times 10^{-1}$). **Case I**.

			$n = 100$				$n = 200$			
			$\rho = 0.25$		$\rho = 0.50$		$\rho = 0.25$		$\rho = 0.50$	
			Empirical	**Smooth**	**Empirical**	**Smooth**	**Empirical**	**Smooth**	**Empirical**	**Smooth**
AL	%RC	t	**Bias(MSE)**	**Bias(MSE)**	**Bias(MSE)**	**Bias(MSE)**	**Bias(MSE)**	**Bias(MSE)**	**Bias(MSE)**	**Bias(MSE)**
Youden Index										
0.75	10	1.0	8.13(1.58)	0.47(0.78)	6.99(1.25)	0.66(0.74)	5.37(0.78)	0.11(0.42)	4.55(0.64)	-0.63(0.40)
0.75	10	1.8	5.80(0.96)	0.36(0.56)	5.16(0.79)	0.04(0.52)	4.02(0.48)	0.33(0.29)	3.62(0.41)	0.26(0.26)
0.75	10	2.9	5.32(0.84)	0.09(0.53)	4.21(0.60)	-0.03(0.45)	3.94(0.46)	0.57(0.28)	3.27(0.34)	0.44(0.22)
1.50	10	1.0	8.37(1.65)	0.28(0.81)	7.12(1.32)	-0.97(0.79)	5.36(0.81)	-0.15(0.46)	4.54(0.66)	-0.94(0.43)
1.50	10	1.8	5.10(0.88)	-0.36(0.56)	4.52(0.72)	-0.62(0.52)	3.31(0.43)	-0.39(0.30)	2.95(0.37)	-0.45(0.27)
1.50	10	2.9	3.52(0.67)	-1.35(0.54)	2.55(0.49)	-1.42(0.46)	2.31(0.34)	-0.84(0.27)	1.69(0.25)	-0.91(0.22)
3.00	10	1.0	8.83(1.84)	0.04(0.90)	7.54(1.52)	-1.19(0.91)	5.69(0.91)	-0.31(0.52)	4.93(0.76)	-1.10(0.49)
3.00	10	1.8	4.68(0.87)	-0.92(0.59)	4.35(0.72)	-0.83(0.55)	2.92(0.42)	-0.89(0.31)	2.84(0.37)	-0.59(0.28)
3.00	10	2.9	2.23(0.62)	-2.27(0.59)	1.72(0.48)	-1.92(0.51)	1.01(0.31)	-1.93(0.31)	0.78(0.23)	-1.61(0.25)
0.75	20	1.0	8.23(1.63)	0.20(0.82)	6.99(1.29)	-0.91(0.78)	5.30(0.78)	-0.17(0.43)	4.41(0.65)	-0.98(0.41)
0.75	20	1.8	5.41(0.93)	-0.11(0.58)	4.84(0.78)	-0.23(0.54)	3.55(0.45)	-0.17(0.30)	3.32(0.39)	-0.06(0.27)
0.75	20	2.9	4.36(0.78)	-0.70(0.56)	3.57(0.56)	-0.53(0.46)	3.01(0.39)	-0.27(0.28)	2.67(0.30)	-0.06(0.22)
1.50	20	1.0	8.49(1.73)	0.05(0.86)	7.16(1.38)	-1.17(0.85)	5.41(0.84)	-0.32(0.48)	4.52(0.69)	-1.18(0.46)
1.50	20	1.8	4.94(0.88)	-0.62(0.58)	4.41(0.73)	-0.71(0.54)	3.06(0.43)	-0.68(0.31)	2.82(0.36)	-0.59(0.27)
1.50	20	2.9	2.91(0.67)	-1.81(0.58)	2.18(0.50)	-1.66(0.49)	1.69(0.32)	-1.39(0.29)	1.32(0.24)	-1.20(0.23)
3.00	20	1.0	9.04(1.96)	0.14(0.97)	7.73(1.62)	-1.32(0.99)	5.89(0.97)	-0.38(0.55)	5.07(0.80)	-1.22(0.53)
3.00	20	1.8	4.74(0.89)	-1.00(0.61)	4.40(0.75)	-0.78(0.58)	2.86(0.43)	-1.00(0.33)	2.86(0.38)	-0.59(0.29)
3.00	20	2.9	1.95(0.64)	-2.48(0.63)	1.56(0.52)	-1.98(0.54)	0.68(0.31)	-2.22(0.33)	0.61(0.24)	-1.71(0.26)
Cut-point										
0.75	10	1.0	-0.30(2.71)	-1.95(1.37)	-0.86(1.72)	-1.94(1.02)	0.03(1.97)	-1.50(0.98)	-0.59(1.18)	-1.53(0.67)
0.75	10	1.8	0.33(1.94)	-1.18(0.87)	-0.51(1.10)	-1.33(0.61)	0.33(1.27)	-0.91(0.60)	-0.42(0.73)	-1.13(0.41)
0.75	10	2.9	0.97(1.70)	-0.49(0.73)	0.46(1.00)	-0.40(0.46)	0.81(1.22)	-0.25(0.53)	0.39(0.66)	-0.34(0.32)
1.50	10	1.0	-0.34(2.85)	-2.21(1.60)	-1.33(1.97)	-2.45(1.30)	-0.28(2.01)	-1.74(1.09)	-0.98(1.32)	-1.92(0.85)
1.50	10	1.8	-0.17(1.81)	-1.62(0.97)	-1.09(1.16)	-1.99(0.80)	-0.14(1.21)	-1.33(0.68)	-1.09(0.81)	-1.82(0.60)
1.50	10	2.9	0.45(1.61)	-1.08(0.76)	-0.45(0.90)	-1.40(0.56)	0.15(1.08)	-0.81(0.51)	-0.63(0.58)	-1.24(0.38)
3.00	10	1.0	-0.82(3.25)	-2.56(2.00)	-1.78(2.38)	-3.00(1.72)	-0.53(2.18)	-2.13(1.30)	-1.4(1.45)	-2.51(1.13)
3.00	10	1.8	-0.75(1.87)	-2.18(1.17)	-1.95(1.39)	-2.85(1.22)	-0.63(1.30)	-1.79(0.83)	-1.91(1.00)	-2.61(0.93)
3.00	10	2.9	-0.47(1.35)	-1.68(0.82)	-1.60(0.95)	-2.45(0.89)	-0.60(0.93)	-1.43(0.56)	-1.61(0.69)	-2.26(0.68)
0.75	20	1.0	-0.15(2.86)	-2.01(1.52)	-1.08(1.77)	-2.09(1.13)	0.05(2.00)	-1.59(1.01)	-0.65(1.23)	-1.72(0.74)
0.75	20	1.8	0.07(1.86)	-1.34(0.92)	-0.72(1.13)	-1.58(0.67)	0.07(1.26)	-1.01(0.61)	-0.69(0.74)	-1.34(0.45)
0.75	20	2.9	0.88(1.70)	-0.70(0.71)	0.23(0.94)	-0.74(0.47)	0.50(1.06)	-0.39(0.49)	0.05(0.61)	-0.64(0.31)
1.50	20	1.0	-0.48(3.12)	-2.30(1.72)	-1.45(2.03)	-2.59(1.41)	-0.33(2.03)	-1.81(1.15)	-1.04(1.39)	-2.09(0.92)
1.50	20	1.8	-0.37(1.75)	-1.74(1.03)	-1.28(1.19)	-2.21(0.89)	-0.34(1.23)	-1.44(0.70)	-1.34(0.84)	-2.03(0.67)
1.50	20	2.9	0.26(1.49)	-1.24(0.77)	-0.77(0.86)	-1.69(0.63)	-0.06(0.99)	-0.95(0.49)	-0.89(0.56)	-1.51(0.43)
3.00	20	1.0	-0.81(3.35)	-2.68(2.13)	-1.94(2.44)	-3.16(1.88)	-0.73(2.20)	-2.22(1.38)	-1.53(1.54)	-2.65(1.20)
3.00	20	1.8	-0.87(1.88)	-2.29(1.24)	-2.16(1.42)	-3.05(1.32)	-0.75(1.25)	-1.92(0.86)	-2.10(1.06)	-2.81(1.03)
3.00	20	2.9	-0.53(1.28)	-1.85(0.86)	-1.88(1.02)	-2.68(1.01)	-0.77(0.90)	-1.55(0.58)	-1.90(0.75)	-2.51(0.78)

Table 4.6: Coverage probabilities (CP) and the average widths (AW) of 95% confidence intervals obtained with the bootstrap using the standard or the percentile method. These results correspond to different sample sizes (n), right censoring rates (%RC), interval lengths (AL) and p values. **Case I,** $t = Q_2 = 1.8$.

			$n = 100$				$n = 200$			
			Standard		**Percentile**		**Standard**		**Percentile**	
p	AL	%RC	CP	AW	CP	AW	CP	AW	CP	AW
0.25	0.75	10	0.934	0.188	0.937	0.187	0.940	0.132	0.937	0.131
0.25	1.50	10	0.938	0.190	0.940	0.189	0.935	0.133	0.935	0.132
0.25	3.00	10	0.942	0.197	0.937	0.196	0.943	0.138	0.941	0.137
0.25	0.75	20	0.926	0.192	0.930	0.191	0.946	0.135	0.950	0.134
0.25	1.50	20	0.932	0.195	0.934	0.193	0.938	0.137	0.931	0.136
0.25	3.00	20	0.941	0.203	0.944	0.201	0.943	0.142	0.938	0.141
0.25	0.75	30	0.937	0.199	0.938	0.197	0.933	0.140	0.931	0.139
0.25	1.50	30	0.938	0.201	0.941	0.200	0.937	0.141	0.936	0.141
0.25	3.00	30	0.936	0.210	0.938	0.209	0.948	0.148	0.944	0.147
0.50	0.75	10	0.928	0.149	0.924	0.148	0.934	0.105	0.940	0.104
0.50	1.50	10	0.929	0.152	0.930	0.151	0.933	0.106	0.935	0.105
0.50	3.00	10	0.938	0.158	0.936	0.157	0.935	0.111	0.934	0.110
0.50	0.75	20	0.914	0.152	0.923	0.151	0.945	0.107	0.944	0.106
0.50	1.50	20	0.924	0.155	0.928	0.154	0.935	0.109	0.937	0.108
0.50	3.00	20	0.933	0.162	0.934	0.161	0.943	0.114	0.944	0.113
0.50	0.75	30	0.924	0.157	0.930	0.156	0.937	0.110	0.933	0.109
0.50	1.50	30	0.929	0.160	0.933	0.158	0.942	0.112	0.944	0.112
0.50	3.00	30	0.927	0.167	0.932	0.166	0.951	0.118	0.953	0.117

4.4 Real data application

In this section, we illustrate the proposed methods using two real-world data sets. The first data set is from the Hypobaric Decompression Sickness study conducted by NASA. The second data set is the Signal Tandmobiel study conducted in the Flanders region of Belgium. The former data consists of subjects either right or interval-censored and hence it corresponds to Case I of our simulation design. Whereas the latter corresponds to Case II of our simulation design and is more general as it contains left, right and interval-censored subjects. The details about these data sets and the results of our analyses are given in the upcoming subsections.

4.4.1 The Hypobaric Decompression Sickness (HDS) data

This data is about a study conducted by NASA from 1983 to 1998 in order to investigate the risk of decompression sickness in hypobaric environments. The out-

come of interest (T) is the time to onset of grade IV venous gas emboli. The data set contains 549 records from 453 males and 96 females who participated in a total of 28 different test procedures. However, some of the subjects underwent in more than one test procedure and hence the total number of individuals actually tested is 238. Among them 169 were right censored and the remaining ones were interval censored; see Conkin and Powell (2001) for more detail. This data is provided in the `R` package `thregI` as the `hdsd` data set (Chen; 2017). Some potential covariates include gender (177 were males), tissue ratio in the 360−minute (TR360) half-time compartment (ranges from 1.04 to 1.84 with median value of 1.771), adynamia status (195 were ambulatory) and age (ranges from 20 to 54 with median value of 30). As in Heagerty and Zheng (2005), we consider a Cox model to derive a risk score from the baseline covariates: X_1 = gender, X_2 = TR360, X_3 = age and X_4 = adynamia. Table 4.7 shows the estimated regression coefficients (β's) with the corresponding $p-$values. Our aim here is to evaluate the predictive performance of the linear score $\hat{\boldsymbol{\beta}}'\boldsymbol{X}$.

The estimated time-dependent AUC and Youden index with their corresponding 95% percentile bootstrap confidence intervals for one, two and three years prediction times are presented in Table 4.8 and 4.9, respectively. The bootstrap confidence intervals were obtained using 1000 bootstrap samples. Because these intervals do not include the null values (0.5 for the AUC and 0 for the Youden index) we can conclude the derived risk score has significant prediction power. Moreover, the estimated values are large indicating good overall classification ability of the risk score. The different estimation methods (based on the AFT, Cox or Spline) give similar results. However, the smoothed ROC based Youden Index estimator results in smaller value than the non-smoothed ROC based estimator. Compared to the proposed methods, the WU method results in smaller estimates and wider confidence intervals. Finally, the Youden index based optimal cut-off point estimates presented in Table 4.10 shows that the nonparametric and AFT methods gives similar results.

Table 4.7: Cox proportional hazards regression estimates for the HDS data

Covariate	Estimate	Std.Error	$p-$values
Age	0.035	0.016	0.028
Sex	0.832	0.402	0.039
TR360	3.015	1.026	0.003
Noadyn	1.425	0.654	0.025

Table 4.8: Estimated AUC with the associated 95% percentile bootstrap confidence intervals (CI) from the HDS data with $t = 1$, 2, and 3 years .

	Proposed			
	Spline	Cox	AFT-Weibull	WU
t	$\widehat{AUC}_t$[95% CI]	$\widehat{AUC}_t$[95% CI]	$\widehat{AUC}_t$[95% CI]	$\widehat{AUC}_t$[95% CI]
1	0.711[0.627, 0.792]	0.724[0.629, 0.792]	0.711[0.627, 0.791]	0.698[0.624, 0.770]
2	0.752[0.676, 0.817]	0.747[0.675, 0.820]	0.752[0.678, 0.818]	0.737[0.665, 0.807]
3	0.736[0.670, 0.801]	0.737[0.669, 0.804]	0.738[0.670, 0.803]	0.733[0.667, 0.803]

Table 4.9: Estimated Youden index with the associated 95% percentile bootstrap confidence intervals (CI) from the HDS data with $t = 1$, 2, and 3 years.

	Non-smooth ROC		Smooth ROC	
	Spline	AFT-Weibull	Spline	AFT-Weibull
t	$\hat{J}_t$[95% CI]	$\hat{J}_t$[95% CI]	$\hat{J}_t$[95% CI]	$\hat{J}_t$[95% CI]
1	0.379[0.299, 0.541]	0.380[0.300, 0.541]	0.288[0.299, 0.540]	0.289[0.301, 0.540]
2	0.502[0.377, 0.621]	0.502[0.378, 0.620]	0.410[0.377, 0.623]	0.411[0.379, 0.623]
3	0.439[0.336, 0.562]	0.440[0.336, 0.562]	0.363[0.337, 0.562]	0.364[0.336, 0.561]

Table 4.10: Estimated Youden index based optimal cut-off values from the HDS data with $t = 1$, 2, and 3 years.

	Non-smooth ROC		Smooth ROC	
	Spline	AFT-Weibull	Spline	AFT-Weibull
1	8.316	8.316	8.093	8.103
2	8.316	8.316	8.342	8.343
3	8.316	8.316	8.300	8.310

4.4.2 The Signal Tandmobiel® (ST) data

The ST data is from a longitudinal prospective study on the oral health condition of primary schoolchildren that was conducted in Flanders region of Belgium for a period of six years (from 1996 to 2001). This study was based on 4430 randomly sampled children who attended the first year of primary school at the beginning of the study. The data set consists of at most six dental observations for each child including the time of tooth emergence, caries experience and data on dietary and oral hygiene habits. Further details on the design of the study can be found in Vanobbergen et al. (2000). This data set is available in the `R` package `bayesSurv`

as the `tandmob2` data set (Komárek; 2018). Here, as the response variable, we consider the time to the emergence of the permanent upper left first premolar (tooth 24 in European dental notation). This data was analyzed in Bogaerts et al. (2017), Gómez et al. (2009) and Fu and Simonoff (2017). The possible predictors of emergence time of the child's tooth include gender (2297 were boys), starting age of brushing teeth (ranges from 1 to 6 with a median value of 2) and BAD which indicates whether a deciduous tooth was removed because of orthodontic reasons or not (215 were removed). Among 4430 subjects, 1612 were right-censored, 42 were left-censored and the rest were interval-censored. As in the above case, we create a risk score for each gender using a Cox model with age and BAD as covariates. The aim here is to compare the predictive performance of these risk scores for boys and girls.

To compare the accuracy of the two risk scores defined above for distinguishing children who develop the permanent upper left first premolar by a given time t and those who develop this after t, we estimated the time-dependent ROC curves and the associated AUC values using the proposed estimators at $t = 6$, and 9 years. Figure 4.2 shows the obtained results. The ROC curves cross each other in most of the cases which complicates the comparison but the AUC helps to overcome this difficulty. When we look at the AUC's and compare the different methods we observe very similar results. This is also the case when we compare girls and boys. We can say that, overall, the discriminatory ability of the risk scores is good but becomes very poor when the time horizon increases from six to nine years.

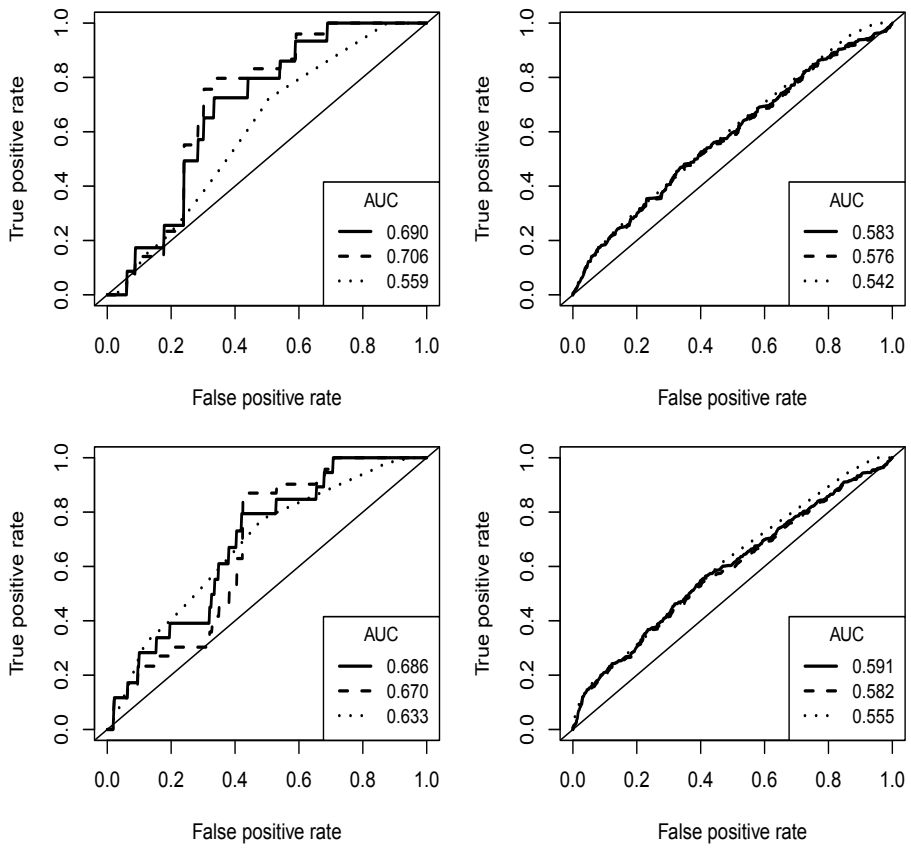

Figure 4.2: The estimated ROC curves and the associated AUC values for boys (top row) and for girls (bottom row) with six years (left panel) and nine years (right panel) prediction time. Spline based method: solid line; AFT-Weibull based method: dashed line; WU method: dotted line.

4.5 Discussion

In this chapter, we proposed and investigated a time-dependent ROC curve and the associated summary indices estimation methods for survival data with more general censorship mechanism. In our proposal, we considered both model-based and nonparametic sieve maximum likelihood approaches in order to estimate the conditional survival function needed for estimating the event status of those subjects who are censored. For the model-based approach both the semiparamtric Cox proportional hazards model and the parametric AFT model were considered.

As for variance estimation and inference method, we proposed a nonparametric bootstrap-based approach for the time-dependent ROC curve and the associated summary measures.

Through simulation study, we see that the proposed, both the model-based and the nonparametric approach, empirical time-dependent ROC curve estimators have better overall finite sample performance than the WU method. Further, when we compare our estimators, the model-based approach has general superiority over the nonparametric method as measured by the MIB and the MISE. Results using the Cox proportional hazards model are similar to the parametric AFT model with slightly better performance from the latter. In addition, the effect of distributional assumption, globally, is minimal as results are very similar. The possible explanation for the less sensitivity of the proposed model-based estimator to misspecification of the parametric form of the underlying population distribution is that its effect is both in the estimation of $\{S(L_i|M_i) - S(t|M_i)\}$ and $\{S(L_i|M_i) - S(R_i|M_i)\}$ and hence it may cancel out to some extent. This robustness is a nice feature because it avoids dealing with the challenge to determine the appropriate parametric form. The general conclusions from the kernel-based smooth ROC estimators were very similar. The smooth ROC estimator, in general, gives better result with respect to MISE than the empirical estimator. In addition, performances of the proposed time-dependent AUC estimators were investigated. Form the simulation study, we observed that both the model-based and the nonparametric AUC estimators have better performance than the WU method as the former have both smaller bias and MSE. In addition, the proposed model-based AUC estimator has better performance than the nonparametric method. Results also revealed that the proposed time-dependent Youden index and the associated cutoff value have satisfactory finite sample performance. To make inference about the ROC curve summary measures, we suggested to use the bootstrap approach. In this chapter, the percentile and normal bootstrap confidence intervals estimation methods were considered. Our simulation results indicate that the former has better performance exhibited by coverage probabilities of desired nominal level and interval widths that are more precise. Our HDS data analysis revealed that the risk score derived by combining the covariates: age, gender, TR360 and adynamia status has a reasonable significant overall accuracy to predict the time to onset of grade IV venous gas emboli. Furthermore, the ST data analysis show that risk score obtained by combining the covariates: BAD and starting age of brushing teeth has moderate overall accuracy to predict teeth emergence time for both boys and girls. The risk score of the boys and girls group has very similar overall predictive ability.

For the model-based approach, we considered both the Cox proportional hazards model and the parametric AFT model to estimate the conditional survival function. This approach is consistent as long as the models are correctly specified. The advantage of this model-based approach is that it can be easily extended to

other survival model classes such as a stratified survival model provided that a consistent estimate for the conditional survival function is obtained. Therefore, extending the proposed methods to accommodate the presence of cure fraction would be an interesting future research topic.

CHAPTER 5

Discussion and Future Research Directions

In modern medical practice, risk scores derived from the statistical models play an important role in making critical decisions, including classifying patients into high and low-risk groups. These attracted many researchers interest to invest considerable time to develop and study predictive performance evaluation methods for the risk scores. The receiver operating characteristic (ROC) curve is the most popular graphical tool to assess the discriminatory ability of the risk scores or markers when predicting a binary disease outcome. The area under the ROC curve (AUC) and Youden index are the two well-known indices to summarize the predictive performance quantitatively. In time-to-event studies, the event status of individuals varies over time, and hence this time-dependent nature should be taken into account. Therefore, to assess the classification accuracy of continuous markers for time-dependent outcomes, various time-dependent extensions of the ROC curve have been proposed. Nonetheless, calculating the time-dependent ROC curve is not straightforward because survival data have several issues, such as the presence of censoring and cure fractions, and hence these need to be addressed. Though various time-dependent ROC curve estimation methods have been proposed in the literature, there is a noticeable gap in the existing literature dealing with the presence of cure fraction, interval censoring, and smoothing the ROC curve. Therefore, this thesis aimed at filling these gaps by proposing some new time-dependent ROC curves and related summary measures estimation and inference methods. This concluding chapter discusses the main contributions of the research works presented

in preceding chapters. Furthermore, this chapter summarizes the main conclusions and provide some suggestions regarding future directions of research related to this work.

5.1 Discussion and general conclusion

Recent technological advancements in medical sciences have stimulated new survival studies, where some patients are long-term survivors, also known as cured or non-susceptibles, especially in oncology, for example, breast cancer. The reason is that the presence of cure fraction in the survival data may produce difficulty to describe it with the classical survival methods. In the past few decades, to address this issue, several methods have been proposed in the literature, including the two most popular modeling frameworks: the mixture cure and the bounded hazards models. Although these methods have been an active area of research for decades and many applications have been published, there is an almost total lack of literature about measuring the predictive ability of these models specifically classification accuracy and the use of these models to derive risk scores. These problems, however, are interrelated, because the latter is in part a consequence of the former, which may explain the reason for the current practice of deriving risk tools from survival models that do not acknowledge the presence of cure.

Motivated by these problems, Chapter 2 intended to make contributions to the existing literature of predictive accuracy measures by addressing two important questions. The first question was to determine the validity of the existing methods: Is it valid to use the time-dependent AUC estimation methods that developed for the classical right censored time-to-event data in the presence of cure fraction? To answer this question, we first derived definitions for cases and controls because the TPR and FPR are defined on the marker values of case and control subjects, respectively. Therefore, for a given time t, it is intuitive to classify subjects as cases if they are susceptible and have survival times less than or equal to t or as controls if they are non-susceptible or have survival times greater than t. These definitions end up with the TPR and FPR definitions that are the same as the one proposed by Heagerty et al. (2000) and hence the existing nonparametric methods such as the simple method of Li et al. (2018) can be used to estimate the time-dependent classification accuracy measures in the presence of cure fraction. In our study, we considered the method of Li et al. (2018) because it uses the available information efficiently, and it outperforms many of other existing methods. This method is based on the well-known Beran estimator, a consistent estimator of survival function in the presence of cure fraction as proved by López-Cheda et al. (2017), and hence its consistency is immanent. The second important question was, what is the loss, in terms of predictive performance, when deriving risk tools from a classical survival model that does not acknowledge the presence of cured fractions? To address these questions, we conducted a comprehensive simulation study, and the

results revealed that the considered time-dependent AUC estimator is consistent for survival data with cure fraction. Furthermore, the results also showed that risk scores derived from the mixture cure model resulted in better predictive performance than the classical Cox proportional hazards model. To further validate this, we analyzed a real-world data from a breast cancer study.

In the statistical literature, many different time-dependent ROC and the related predictive accuracy measures methods for the right censored time-to-event have been proposed. Although it is well-documented that, in the case of uncensored data, a smoothed estimator has better performance than an empirical estimator, almost all of the existing time-dependent ROC curve estimation methods are empirical. To the best of our knowledge, Martínez-Camblor and Pardo-Fernández (2018) is the only smooth estimator. To address this scarcity, in Chapter 3, we proposed some nonparametric time-dependent ROC curve estimators for right censored survival data. To develop these estimators, using the ROC definition and borrowing the idea of defining an event state from Li et al. (2018) and Martínez-Camblor et al. (2016), we first formulated the ROC as a weighted empirical distribution function. The weights are nothing but the conditional survival probabilities given the observed data, which we suggested to estimate using the Beran estimator as in Li et al. (2018). To develop a smooth ROC estimator, we replaced the indicator function in the empirical ROC estimator with a kernel distribution function. One major challenge in this method is choosing the bandwidth parameters need for calculating the Beran weights and for smoothing the ROC curve. For the former, relying on the finding and justifications of Li et al. (2018), we suggested to use the bandwidth selection methods proposed for the classical (uncensored) kernel density estimator. To select the bandwidth parameter necessitated to smooth the ROC, we proposed three data-driven methods: the normal reference, the plug-in, and the cross-validation. The first two methods are based on the asymptotic results of the proposed smoothed ROC estimator (see Section 3.2.3), whereas the last one is a direct extension of the method proposed by Bowman et al. (1998). The additional challenge of the proposed method is a boundary problem that arises from the boundedness of the support of the random variable of interest. In order to overcome this problem, we proposed an alternative estimator, using a basic probability theory, based on a quantile transformation method, specifically the probit transformation. This transformation aimed to make the support of the random variable of interest unbounded by sending away the boundaries to $\pm\infty$, and hence the kernel smoothed time-dependent ROC estimator of this transformed random variable become free of the boundary bias. As for variance estimation and confidence interval construction, we introduced a data-based nonparametric bootstrap approach. Through an extensive simulation study, the proposed smoothed time-dependent ROC with boundary correction is shown to perform better than the ROC estimator without boundary correction. Furthermore, the former is relatively quite robust to the choice of bandwidth parameter needed to smooth the ROC curve, which is

another advantage of correcting the boundary problem. In addition, the proposed bandwidth selection methods produce very similar results. The simulation results also revealed that the proposed smoothed ROC estimator, in general, outperforms both the empirical (non-smoothed) estimator and the existing competitor (i.e. the MP) method. The bootstrap confidence interval estimation methods considered produced coverage probabilities that are reasonably close to the nominal level. Finally, the application of the proposed methods is illustrated using the well-known Mayo PBC data. The proposed methods are implemented in the open-source R package `cenROC`, which is publicly available from the Comprehensive R Archive Network (CRAN) site (`https://CRAN.Rproject.org/package=cenROC`). The usage of this package and the functions therein are discussed in Appendix A.

In many practical situations, interval censoring is a common phenomenon when the event of interest only known to have occurred within a given time interval. Though several interval censored data analysis methods have been proposed and studied over the past few decades, the literature about models predictive accuracy estimation methods is rather scarce compared to the right censored data. Chapter 4 addressed this literature gap by introducing some new time-dependent ROC curve and related summary measures estimation methods that take the interval censored nature into account. The event status of those subjects whose censoring interval includes the prediction time t is not known, and hence we replaced it with a value obtained from the estimated conditional survival probability given the marker. For the estimation of the conditional survival probabilities, we proposed parametric, semiparametric and nonparametric approaches. The parametric and semiparametric methods are, of course, the Cox proportional hazards and the accelerated failure time (AFT) models, respectively. We used a spline-based sieve maximum likelihood estimation approach proposed by Wu and Zhang (2012); Wu et al. (2020), the detail about this method is presented in Section 1.2.1, to estimate the conditional survival probabilities nonparametrically. Solving the problem of identifying subjects event status in this way, we first developed the time-dependent ROC curve estimators as a weight empirical distribution function and then, relying on the results from the third chapter, proposed a boundary corrected smoothed ROC estimator. As for the summary measure, we considered the two well-known methods: the time-dependent area under the ROC curve (AUC) and the time-dependent Youden index. The latter, however, not only summarizes the classification accuracy of a marker but also provides an optimal cut-point for future diagnosis. We suggested a method using bootstrapping to estimate standard error and calculate confidence intervals. To achieve the latter, we considered the standard normal and the percentile interval estimation methods. A comprehensive simulation study was conducted to assess the finite-sample performance of the proposed methods, which demonstrated the superiority of our methods over the considered competitor. The model-based approaches produced comparable performance, with the AFT having a slight advantage. Furthermore, the smoothed estimators generally performs bet-

ter, in terms of the mean integrated squared errors, than the empirical methods. The percentile bootstrap confidence interval has a better performance than the standard normal as measured by coverage probability and average width. Finally, the proposed methods are illustrated using two real-life data sets: one from the Hypobaric Decompression Sickness study conducted by NASA and the other from the Signal Tandmobiel study conducted in the Flanders region of Belgium. These methods are also implemented in the R package `cenROC`.

5.2 Future research directions

The research work presented in this thesis opens several fascinating future research opportunities. Some of these are direct extensions of the proposed approaches, while others will need combining with other research findings. In this section, we discuss some of these future research perspectives.

In this thesis, to summarize the time-dependent ROC curve quantitatively, we used the time-dependent AUC. This method is the most popular and well developed overall classification accuracy measure with meaningful interpretations. Nevertheless, it summarizes the entire ROC curve, including parts that usually are not relevant for practical applications. In population screening, high FPR can cause high financial costs; hence the region of the ROC curve that corresponds to the high FPR values is not the primary concern (Dodd and Pepe; 2003). Also, when ROC curves of two markers cross, the AUC may not be the best performance indicator because the ROC curve with smaller AUC may have higher TPR in the clinically relevant range of FPR values. Therefore, it is interesting to consider alternative ROC curve summary measures that can overcome this problem. One of the solutions to this problem is the partial area under the ROC curve (pAUC), which summarizes the region of the ROC curve that corresponds to the prespecified range of clinically relevant FPR values. Given a range of time-dependent FPR, $(0, u_1)$, the time-dependent pAUC is defined as an integral of the time-dependent ROC function over the given range, that is

$$pAUC_t = \int_0^{u_1} ROC_t(s)ds.$$

When $u_1 = 1$, the $pAUC_t$ reduces to the classical AUC_t. The partial AUC has some interesting interpretations including a modified form of the probability interpretation of Hanley and McNeil (1982); for more detail, see Dodd and Pepe (2003). Another natural summary measure for the ROC curve is to use, for clinical reasons, a point on it rather than a range of values. To this end, one can either compute the TPR for a given FPR or the FPR for a given TPR. The former is just a point on the ROC curve whereas the latter one can be considered as a point on the inverse

of the ROC function and mathematically defined as follows:

$$\mathrm{ROC}_t^{-1}(p) = \mathrm{FPR}_t(\mathrm{TPR}_t^{-1}(p)),$$

where $p \in (0,1)$ and $m = \mathrm{TPR}_t^{-1}(p)$ is the threshold value such that $p = \mathrm{TPR}_t(m)$. Using the time-dependent TPR and FPR estimators given in Section 2.2.1, we can readily obtain the estimate of this function.

Another interesting summary measure to consider is the C-index, which is a very widely used summary index to measure the overall discriminatory ability of a prognostic marker. In the traditional binary outcome data, it is the same as the AUC. In contrast to this, in the case of time-to-event data, they are not the same, but they are related. The advantage of the C-index is that, unlike the time-dependent AUC, it is time-independent. The C-index is defined as

$$C = \int_t AUC(t)w(t)dt,$$

where $w(t)$ is a weight function and $AUC(t)$ is the time-dependent AUC at the prediction time t. The two commonly used weight functions are $2f(t)S(t)$ and $f(t)$, where $f(.)$ and $S(.)$ are the density and survival function of the event time, respectively. In the former case (i.e., $w(t) = 2f(t)S(t)$), $C = P(M_i > M_j | T_i < T_j) = E_T(AUC(T) \times 2 \times S(T))$ and when $w(t) = f(t)$, the C-index is equals to the expected value of the time-dependent AUC, i.e., $C = \int_t AUC(t)f(t) = E_T(AUC(T))$ (Heagerty and Zheng; 2005; Schmid et al.; 2013). Using the proposed time-dependent AUC estimator together with some existing approaches for estimating the survival function (see, for example, Li and Ma (2011); Schmid et al. (2013)), the C-index estimate can be readily obtained.

This thesis mainly focused on the sensitivity and specificity based classification accuracy measures. These measures are crucial to address a question related to the discriminatory ability of a diagnostic test or marker. Therefore, they are mainly concerned about assessing a diagnostic tool. In clinical medicine, however, assessing individuals' health status given the diagnostic test result also has great importance, since the final goal of any health care service is to help patients. Therefore, it is necessary to study measures that are relevant to address this problem. To this end, the positive predictive value (PPV) and negative predictive value (NPV) are essential measures for categorizing subjects as having and not having an event of interest. The PPV is the probability that the disease is present, given a positive test result, and the NPV is the probability that the disease is absent, given a negative test result. These are given by

$$PPV(m) = P(T \leq t | M > m) = \frac{P(M > m | T \leq t)P(T \leq t)}{P(M > m)},$$

and

$$NPV(m) = P(T > t|M \leq m) = \frac{P(M \leq m|T > t)P(T > t)}{P(M \leq m)},$$

respectively. These quantities can be easily estimated using the TPR, FPR, and the weight estimator developed in Section 2.2.1.

In Chapter 2, our main interest was to assess the markers ability to discriminate subjects with and without the event of interest at a given prediction time t. However, time-to-event data with cure fractions analysis have an additional attractive objective that is modeling the cure status. Therefore, evaluating the ability of a risk score or marker to discriminate subjects who are cured and who are not cured is of interest. To this end, Amico and Van Keilegom (2018a) proposed a ROC curve estimation method based on a mixture cure model. The main drawback of this approach is that it may suffer from model misspecification and thus may result in biased estimates. Therefore, it would be interesting to consider a fully nonparametric ROC curve and the corresponding summary measures estimation approach. To this end, combining the finding of López-Cheda et al. (2017) and Beyene et al. (2019) with the proposals of Amico and Van Keilegom (2018a), the TPR and FPR estimators can be defined as

$$\widehat{TPR}(m) = \hat{\mathrm{P}}(M > m|T = \infty) = \frac{\sum_i \hat{W}_{1i} I(M_i > m)}{\sum_i \hat{W}_{1i}},$$

$$\widehat{FPR}(m) = \hat{\mathrm{P}}(M > m|T < \infty) = \frac{\sum_i (1 - \hat{W}_{1i}) I(M_i > m)}{\sum_i (1 - \hat{W}_{1i})},$$

where $\hat{W}_{1i}$ is the cure status estimate of the i^{th} subject that is defined as

$$\hat{W}_{1i} = I(Y_i > \tau) + (1 - \Delta_i) I(Y_i \leq \tau) \frac{1 - \hat{p}(M_i)}{\hat{S}(Y_i|M_i)},$$

where τ is the cure threshold and it is suggested to consider the largest uncensored observed time $Y^1_{\max}$, $1 - \hat{p}(M_i)$ is the cure probability estimate, and $\hat{S}(Y_i|M_i)$ is the population survival function estimate. Following López-Cheda et al. (2017), both the cure probability and the population survival function can be estimated using the classical Beran estimator. For more detail about the nonparametric survival function estimation in the presence of cure fraction, see López-Cheda et al. (2017).

In Chapter 3, to overcome the boundary problem, we used a quantile transformation that transforms the random variable with bounded unit interval $[0, 1]$ support to the real number. This method is intuitive because it is based on the basic property of probability. Though this approach has improved the performance of the proposed kernel smoothed ROC curve estimator, considering other existing boundary correction methods and comparing their performance with the quantile

transformation technique would be an interesting research problem. In the literature, there are some alternative boundary correction methods. The generalized reflection is one of the popular methods for correcting the boundary problem of the kernel distribution estimator. Koláček and Karunamuni (2009) proposed this method to remove the bias near or at the boundary points of the kernel smoothed ROC curve for the uncensored data. Alternatively, one could also consider the boundary-modified kernel distribution function approach proposed to generate a smoothed distribution function for complete data (Tenreiro; 2013, 2018).

The main challenge in the kernel smoothing approach is selecting the bandwidth parameter optimally. To this end, in Chapter 3, we proposed data-driven bandwidth selection methods. However, the proposed plug-in method requires choosing an additional bandwidth parameter needed to estimate $\kappa(ROC)$. To overcome this difficulty, we used the normal-reference bandwidth selection method given in equation (3.10). This approach, however, may not be optimal, at least asymptotically, and hence developing the asymptotic theory of $\hat{E}\Big\{\widehat{\hat{\mathcal{W}}ROC^{(3)}}(Z)\Big\}$ and selecting the smoothing parameter based on this would be an exciting but very challenging future work. This asymptotic theory will also help to extend the multistage plug-in method of Polansky and Baker (2000) for the case of weighted data. In the literature, the other existing bandwidth selection approach is the direct plug-in method proposed Altman and Léger (1995), extending this method to account for the weight effect, and assessing its performance would also be a challenging problem.

References

Altman, N. and Léger, C. (1995). Bandwidth selection for kernel distribution function estimation, *Journal of Statistical Planning and Inference* **46**(2): 195 – 214.

Amico, M. and Van Keilegom, I. (2018a). Assessing cure status prediction from survival data using ROC curves, *FEB Research Report KBI_1827* .

Amico, M. and Van Keilegom, I. (2018b). Cure models in survival analysis, *Annual Review of Statistics and Its Application* **5**(1): 311–342.

Amico, M., Van Keilegom, I. and Legrand, C. (2019). The single-index/Cox mixture cure model, *Biometrics* **75**(2): 452–462.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study, *The Annals of Statistics* **10**(4): 1100–1120.

Anderson-Bergman, C. (2017a). An efficient implementation of the EMICM algorithm for the interval censored NPMLE, *Journal of Computational and Graphical Statistics* **26**(2): 463–467.

Anderson-Bergman, C. (2017b). icenReg: Regression models for interval censored data in R, *Journal of Statistical Software* **81**(12): 1–23.

Aoki, K., Misumi, J., Kimura, T., Zhao, W. and Xie, T. (1997). Evaluation of cutoff levels for screening of gastric cancer using serum pepsinogens and distributions of levels of serum pepsinogen i, ii and of pg i / pg ii ratios in a gastric cancer case-control study, *Journal of Epidemiology* **7**(3): 143–151.

Baršić, B., Marton, E., Himbele, J. and Ravlić, Ž. (1996). Evaluation of the Glasgow coma scale score in critically ill infectious disease patients, *Infection* **24**(4): 297–300.

Beran, R. (1981). Nonparametric regression with randomly censored survival data, *Technical report*, Univ. California, Berkeley.

Beyene, K. M. and El Ghouch, A. (2020a). *cenROC: Estimation of the time-dependent ROC curve and AUC for censored survival data*. R package version 1.2.0.

Beyene, K. M. and El Ghouch, A. (2020b). Smoothed time-dependent receiver operating characteristic curve for right censored survival data, *Statistics in Medicine* pp. 1–24.

Beyene, K. M. and El Ghouch, A. (2020c). Time-dependent ROC curve estimation for interval censored data, *[submitted]* pp. 1–20.

Beyene, K. M., El Ghouch, A. and Oulhaj, A. (2019). On the validity of time-dependent AUC estimation in the presence of cure fraction, *Biometrical Journal* **61**(6): 1430–1447.

Blanche, P., Dartigues, J.-F. and Jacqmin-Gadda, H. (2013). Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring, *Biometrical Journal* **55**(5): 687–704.

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy, *Journal of the Royal Statistical Society. Series B (Methodological)* **11**(1): 15–53.

Bogaerts, K., Komarek, A. and Lesaffre, E. (2017). *Survival analysis with interval-censored data: A practical approach with examples in R, SAS, and BUGS*, Chapman and Hall/CRC.

Bowman, A., Hall, P. and Trvan, T. (1998). Bandwidth selection for the smoothing of distribution functions, *Biometrika* **85**: 799 – 808.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship, *The Annals of Statistics* **2**(3): 437–453.

Cai, T. and Cheng, S. (2008). Robust combination of multiple diagnostic tests for classifying censored event times, *Biostatistics* **9**(2): 216–233.

Cai, T., Pepe, M. S., Zheng, Y., Lumley, T. and Jenny, N. S. (2006). The sensitivity and specificity of markers for event times, *Biostatistics* **7**(2): 182–197.

Chambless, L. E. and Diao, G. (2006). Estimation of time-dependent area under the ROC curve for long-term risk prediction, *Statistics in Medicine* **25**(20): 3474–3486.

Chen, D.-G. D., Sun, J. and Peace, K. E. (2012). *Interval-censored time-to-event data: methods and applications*, Chapman and Hall/CRC.

Chen, M.-H. (2017). *thregI: Threshold regression for interval-censored data with a cure rate option*. R package version 1.0.4.

Chen, M.-H., Ibrahim, J. G. and Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction, *Journal of the American Statistical Association* **94**(447): 909–919.

Collett, D. (2015). *Modelling survival data in medical research*, Chapman and Hall/CRC.

Conkin, J. and Powell, M. (2001). Lower body adynamia as a factor to reduce the risk of hypobaric decompression sickness., *Aviation, Space, and Environmental Medicine* **72**(3): 202.

Cook, N. R. (2008). Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve, *Clinical Chemistry* **54**(1): 17–23.

Cox, D. R. (1972). Regression models and life-tables, *Journal of the Royal Statistical Society. Series B (Methodological)* **34**(2): 187–220.

Dabrowska, D. M. (1987). Non-parametric regression with censored survival time data, *Scandinavian Journal of Statistics* **14**(3): 181–197.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate, *The Annals of Statistics* **17**(3): 1157–1167.

Dehghan, M. H. and Duchesne, T. (2011a). A generalization of Turnbull's estimator for nonparametric estimation of the conditional survival function with interval-censored data, *Lifetime Data Analysis* **17**(2): 234–255.

Dehghan, M. H. and Duchesne, T. (2011b). On the performance of some non-parametric estimators of the conditional survival function with interval-censored data, *Computational Statistics & Data Analysis* **55**(12): 3355 – 3364.

Díaz-Coto, S. (2018). *smoothROCtime: Smooth time-dependent ROC curve estimation*. R package version 0.1.0.

Díaz-Coto, S., Martínez-Camblor, P. and Corral-Blanco, N. O. (2020). Cumulative/dynamic ROC curve estimation under interval censorship, *Journal of Statistical Computation and Simulation* **90**(9): 1570–1590.

Diciccio, T. J. and Romano, J. P. (1988). A review of bootstrap confidence intervals, *Journal of the Royal Statistical Society. Series B (Methodological)* **50**(3): 338–354.

Dodd, L. (2001). Regression methods for areas and partial areas under the ROC curve, *Ph. D. thesis, University of Washington* .

Dodd, L. E. and Pepe, M. S. (2003). Partial AUC estimation and regression, *Biometrics* **59**(3): 614–623.

Edmonson, J. H., Fleming, T. R., Decker, D., Malkasian, G., Jorgensen, E., Jefferies, J., Webb, M. and Kvols, L. (1979). Different chemotherapeutic sensitivities and host factors affecting prognosis in advanced ovarian carcinoma versus minimal residual disease., *Cancer treatment reports* **63**(2): 241–247.

Efron, B. (1979). Bootstrap methods: another look at the Jackknife, *The Annals of Statistics* **7**(1): 1–26.

Efron, B. (1981). Nonparametric standard errors and confidence intervals, *Canadian Journal of Statistics* **9**(2): 139–158.

Egan, J. P. (1975). *Signal detection theory and ROC analysis*, Academic press.

Etzioni, R., Pepe, M., Longton, G., Hu, C. and Goodman, G. (1999). Incorporating the time dimension in receiver operating characteristic curves: A case study of prostate cancer, *Medical Decision Making* **19**(3): 242 – 251.

Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions, *Statistica Neerlandica* **37**(2): 73–83.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data, *Biometrics* **42**(4): 845–854.

Fluss, R., Faraggi, D. and Reiser, B. (2005). Estimation of the Youden index and its associated cutoff point, *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **47**(4): 458–472.

Fu, W. and Simonoff, J. S. (2017). Survival trees for interval-censored survival data, *Statistics in Medicine* **36**(30): 4831–4842.

Geenens, G. (2014). Probit transformation for kernel density estimation on the unit interval, *Journal of the American Statistical Association* **109**(505): 346–358.

Gómez, G., Calle, M. L., Oller, R. and Langohr, K. (2009). Tutorial on methods for interval-censored data and their implementation in R, *Statistical Modelling* **9**(4): 259–297.

González-Manteiga, W. and Cadarso-Suárez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications, *Journal of Nonparametric Statistics* **4**(1): 65–78.

Green, D. M., Swets, J. A. et al. (1966). *Signal detection theory and psychophysics*, Vol. 1, Wiley New York.

Greiner, M., Pfeiffer, D. and Smith, R. (2000). Principles and practical application of the receiver operating characteristic analysis for diagnostic tests, *Preventive Veterinary Medicine* **45**(1-2): 23–41.

Hall, P., Hyndman, R. J. and Fan, Y. (2004). Nonparametric confidence intervals for receiver operating characteristic curves, *Biometrika* **91**(3): 743–750.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve., *Radiology* **143**(1): 29–36.

Heagerty, P. J., Lumley, T. and Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker, *Biometrics* **56**(2): 337–344.

Heagerty, P. J. and Saha-Chaudhuri, P. (2013). *survivalROC: Time-dependent ROC curve estimation from censored survival data*. R package version 1.0.3.

Heagerty, P. J. and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves, *Biometrics* **61**(1): 92–105.

Hilden, J. and Glasziou, P. P. (1996). Regret graphs, diagnostic uncertainty and Youden's index, *Statistics in Medicine* **15**(10): 969–986.

Horowitz, J. (2001). *Chapter 52 The Bootstrap*, Handbook of Econometrics, Elsevier, pp. 3159–3228.

Hsieh, F. and Turnbull, B. W. (1996). Nonparametric methods for evaluating diagnostic tests, *Statistica Sinica* **6**: 47–62.

Hung, H. and Chiang, C.-T. (2010). Estimation methods for time-dependent AUC models with survival data, *Canadian Journal of Statistics* **38**(1): 8–26.

Hwang, Y.-T., Hung, Y.-H., Wang, C. C. and Terng, H.-J. (2018). Finding the optimal threshold of a parametric ROC curve under a continuous diagnostic measurement, *REVSTAT–Statistical Journal* **16**(1): 23–43.

Jones, M. C. (1990). The performance of kernel density functions in kernel distribution function estimation, *Statistics & Probability Letters* **9**(2): 129–132.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, Vol. 360, John Wiley & Sons.

Kamarudin, A. N., Cox, T. and Kolamunnage-Dona, R. (2017). Time-dependent ROC curve analysis in medical research: current methods and applications, *BMC Medical Research Methodology* **17**(1): 1–19.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* **53**(282): 457–481.

Karunamuni, R. J. and Alberts, T. (2005). On boundary correction in kernel density estimation, *Statistical Methodology* **2**(3): 191–212.

Kim, J. S. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(2): 489–502.

Koláček, J. and Karunamuni, R. J. (2009). On boundary correction in kernel estimation of ROC curves, *Austrian Journal of Statistics* **38**(1): 17–32.

Komárek, A. (2018). *bayesSurv: Bayesian survival regression with flexible error and random effects distributions*. R package version 3.2.

Kreager, P. (1988). New light on graunt, *Population Studies* **42**(1): 129–140.

Krzanowski, W. J. and Hand, D. J. (2009). *ROC curves for continuous data*, 1st edn, Chapman and Hall.

Lasko, T. A., Bhagwat, J. G., Zou, K. H. and Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics, *Journal of Biomedical Informatics* **38**(5): 404 – 415.

Law, C. G. and Brookmeyer, R. (1992). Effects of mid-point imputation on the analysis of doubly censored data, *Statistics in Medicine* **11**(12): 1569–1578.

Li, J. and Ma, S. (2011). Time-dependent ROC analysis under diverse censoring patterns, *Statistics in Medicine* **30**(11): 1266–1277.

Li, L., Greene, T. and Hu, B. (2018). A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data, *Statistical Methods in Medical Research* **27**(8): 2264–2278.

Li, L. and Wu, C. (2016). *tdROC: Nonparametric estimation of time-dependent ROC curve from right censored survival data*. R package version 1.0.

Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics: theory and practice*, Princeton University Press.

Lin, J., Wu, Y., Wang, X. and Owzar., K. (2018). *intcensROC: Fast spline function based constrained maximum likelihood estimator for AUC estimation of interval censored survival data*. R package version 0.1.1.

Liu, X. (2012). Classification accuracy and cut point selection, *Statistics in Medicine* **31**(23): 2676–2686.

Liu, X. and Jin, Z. (2015). Optimal survival time related cut point with censored data, *Statistics in Medicine* **34**(3): 515–524.

Lloyd, C. J. and Yong, Z. (1999). Kernel estimators of the ROC curve are better than empirical, *Statistics & Probability Letters* **44**(3): 221 – 228.

López-Cheda, A., Cao, R., Jácome, M. A. and Van Keilegom, I. (2017). Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models, *Computational Statistics & Data Analysis* **100**(105): 144–165.

López-Ratón, M., Rodríguez-Álvarez, M. X., Cadarso-Suárez, C. and Gude-Sampedro, F. (2014). Optimalcutpoints: An R package for selecting optimal cutpoints in diagnostic tests, *Journal of Statistical Software* **61**(8): 1–36.

Lusted, L. B. (1960). Logical analysis in roentgen diagnosis, *Radiology* **74**(2): 178–193.

Marron, J. S. and Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation, *Journal of the Royal Statistical Society. Series B (Methodological)* **56**(4): 653–671.

Martínez-Camblor, P., Bayón, G. F. and Pérez-Fernández, S. (2016). Cumulative/dynamic ROC curve estimation, *Journal of Statistical Computation and Simulation* **86**(17): 3582 – 3594.

Martínez-Camblor, P. and Pardo-Fernández, J. C. (2018). Smooth time-dependent receiver operating characteristic curve estimators, *Statistical Methods in Medical Research* **27**(3): 651 – 674.

McIntosh, M. W. and Pepe, M. S. (2002). Combining several screening tests: Optimality of the risk score, *Biometrics* **58**(3): 657–664.

Meier, P. (1975). Estimation of a distribution function from incomplete observations, *Journal of Applied Probability* **12**(S1): 67–87.

Metz, C. E. (1989). Some practical issues of experimental design and data analysis in radiological ROC studies., *Investigative Radiology* **24**(3): 234–245.

Nadaraya, E. (1965). On non-parametric estimates of density functions and regression curves, *Theory of Probability & Its Applications* **10**(1): 186–190.

Nadaraya, E. A. (1964). Some new estimates for distribution functions, *Theory of Probability & Its Applications* **9**(3): 497–500.

Pan, W. (1999). Extending the iterative convex minorant algorithm to the Cox model for interval-censored data, *Journal of Computational and Graphical Statistics* **8**(1): 109–120.

Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*, Oxford Statistical Sciences Series.

Pepe, M. S., Zheng, Y., Jin, Y., Huang, Y., Parikh, C. R. and Levy, W. C. (2008). Evaluating the ROC performance of markers for future events, *Lifetime Data Analysis* **14**(1): 86–113.

Perkins, N., J. and Schisterman, E., F. (2006). The inconsistency of optimal cut-points using two ROC based criteria, *American Journal of Epidemiology* **163**: 670–675.

Polansky, A. M. and Baker, E. R. (2000). Multistage plugin bandwidth selection for kernel distribution function estimates, *Journal of Statistical Computation and Simulation* **65**(1-4): 63–80.

Portier, F., El Ghouch, A. and Van Keilegom, I. (2017). Efficiency and bootstrap in the promotion time cure model, *Bernoulli* **23**(4B): 3437–3468.

Pulit, M. (2016). A new method of kernel-smoothing estimation of the ROC curve, *Metrika* **79**(5): 603–634.

Qin, J. and Zhang, B. (2010). Best combination of multiple diagnostic tests for screening purposes, *Statistics in Medicine* **29**(28): 2905–2919.

R Core Team (2020). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

Reiss, R.-D. (1981). Nonparametric estimation of smooth distribution functions, *Scandinavian Journal of Statistics* **8**(2): 116–119.

Rota, M., Antolini, L. and Valsecchi, M. G. (2015). Optimal cut-point definition in biomarkers: the case of censored failure time outcome, *BMC Medical Research Methodology* **15**(1): 1–11.

Ruopp, M. D., Perkins, N. J., Whitcomb, B. W. and Schisterman, E. F. (2008). Youden index and optimal cut-point estimated from observations affected by a lower limit of detection, *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **50**(3): 419–430.

Saha-Chaudhuri, P. and Heagerty, P. (2013). Non-parametric estimation of a time-dependent predictive accuracy curve, *Biostatistics* **14**(1): 42–59.

Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions, *Journal of Statistical Planning and Inference* **35**(1): 65 – 75.

Schmid, M., Kestler, H. A. and Potapov, S. (2013). On the validity of time-dependent AUC estimators, *Briefings in Bioinformatics* **16**(1): 153–168.

Schmid, M. and Potapov, S. (2012). A comparison of estimators to evaluate the discriminatory power of time-to-event models, *Statistics in Medicine* **31**(23): 2588–2609.

Shapiro, D. E. (1999). The interpretation of diagnostic tests, *Statistical Methods in Medical Research* **8**(2): 113–134.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society. Series B (Methodological)* **53**(3): 683–690.

Shen, W., Ning, J. and Yuan, Y. (2015). A direct method to evaluate the time-dependent predictive accuracy for biomarkers, *Biometrics* **71**(2): 439–449.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Chapman and Hall, London.

Sima, C. S. and Gönen, M. (2013). Optimal cutpoint estimation with censored data, *Journal of Statistical Theory and Practice* **7**(2): 345–359.

Slate, E. H. and Turnbull, B. W. (2000). Statistical models for longitudinal biomarkers of disease onset, *Statistics in Medicine* **19**(4): 617 – 637.

Spackman, K. A. (1989). Signal detection theory: Valuable tools for evaluating inductive learning, *Proceedings of the sixth international workshop on Machine learning*, pp. 160–163.

Stevens, R. J., Kothari, V., Adler, A. I., Stratton, I. M. and Holman, R. R. (2001). The UKPDS risk engine: a model for the risk of coronary heart disease in type ii diabetes (UKPDS 56), *Clinical Science* **101**(6): 671–679.

Sun, J. (2007). *The statistical analysis of interval-censored failure time data*, Springer.

Sun, X. and Chen, C. (2010). Comparison of Finkelstein's method with the conventional approach for interval-censored data analysis, *Statistics in Biopharmaceutical Research* **2**(1): 97–108.

Swanepoel, J. W. (1988). Mean intergrated squared error properties and optimal kernels when estimating a diatribution function, *Communications in Statistics - Theory and Methods* **17**(11): 3785–3799.

Swets, J. A. (1973). The relative operating characteristic in psychology, *Science* **182**(4116): 990–1000.

Swets, J. and Pickett, R. (1982). *Evaluation of diagnostic systems: methods from signal detection theory*, Academic Press.

Tableman, M. and Kim, J. S. (2003). *Survival analysis using S: analysis of time-to-event data*, Chapman and Hall/CRC.

Taweab, F. and Ibrahim, N. A. (2014). Cure rate models: a review of recent progress with a study of change-point cure models when cured is partially known, *Journal of Applied Sciences* **14**(7): 609–616.

Tenreiro, C. (2013). Boundary kernels for distribution function estimation, *REVSTAT Statistical Journal* **11**(2): 169–190.

Tenreiro, C. (2018). A new class of boundary kernels for distribution function estimation, *Communications in Statistics-Theory and Methods* **47**(21): 5319–5332.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model, *The Annals of Statistics* **9**(1): 93–108.

Tsodikov, A. (1998a). Asymptotic efficiency of a proportional hazards model with cure, *Statistics & Probability Letters* **39**(3): 237–244.

Tsodikov, A. (1998b). A proportional hazards model taking account of long-term survivors, *Biometrics* **54**(4): 1508–1516.

Tsodikov, A. (2001). Estimation of survival based on proportional hazards when cure is a possibility, *Mathematical and Computer Modelling* **33**(12): 1227–1236.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data, *Journal of the Royal Statistical Society. Series B (Methodological)* **38**(3): 290–295.

Uno, H., Cai, T., Tian, L. and Wei, L. (2007). Evaluating prediction rules for t-year survivors with censored regression models, *Journal of the American Statistical Association* **102**(478): 527–537.

Van Calster, B., McLernon, D. J., Van Smeden, M., Wynants, L. and Steyerberg, E. W. (2019). Calibration: the Achilles heel of predictive analytics, *BMC Medicine* **17**(1): 1–7.

Van Keilegom, I. and Akritas, M. G. (1999). Transfer of tail information in censored regression models, *Annals of Statistics* **27**(5): 1745–1784.

Van Keilegom, I. and Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression, *Annals of the Institute of Statistical Mathematics* **49**(3): 467–491.

Vanobbergen, J., Martens, L., Lesaffre, E. and Declerck, D. (2000). The signal-tandmobiel project a longitudinal intervention health promotion study in flanders (belgium): baseline and first year results, *European Journal of Paediatric Dentistry* **2**: 87–96.

Wand, M. P., Marron, J. S. and Ruppert, D. (1991). Transformations in density estimation, *Journal of the American Statistical Association* **86**(414): 343–353.

Wang, Y. (2008). Dimension-reduced nonparametric maximum likelihood computation for interval-censored data, *Computational Statistics & Data Analysis* **52**(5): 2388–2402.

Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., van Gelder, M. E. M., Yu, J., Jatkoe, T., Berns, E. M., Atkins, D. and Foekens, J. A. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer, *The Lancet* **365**(9460): 671–679.

Watson, G. S. (1964). Smooth regression analysis, *Sankhyā: The Indian Journal of Statistics, Series A* **26**(4): 359–372.

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data, *Journal of the American Statistical Association* **92**(439): 945–959.

Wen, K. and Wu, X. (2015). An improved transformation-based kernel estimator of densities on the unit interval, *Journal of the American Statistical Association* **110**(510): 773–783.

Wu, Y. and Cook, R. J. (2020). Assessing the accuracy of predictive models with interval-censored data, *Biostatistics* .

Wu, Y., Wang, X., Lin, J., Jia, B. and Owzar, K. (2020). Predictive accuracy of markers or risk scores for interval censored survival data, *Statistics in Medicine* **39**(18): 2437–2446.

Wu, Y. and Zhang, Y. (2012). Partially monotone tensor spline estimation of the joint distribution function with bivariate current status data, *The Annals of Statistics* **40**(3): 1609–1636.

Xu, J. and Peng, Y. (2014). Nonparametric cure rate estimation with covariates, *Canadian Journal of Statistics* **42**(1): 1–17.

Yakovlev, A. Y., Asselain, B., Bardou, V., Fourquet, A., Hoang, T., Rochefediere, A. and Tsodikov, A. (1993). A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer, *Biometrie et Analyse de Donnees Spatio-Temporelles* **12**: 66–82.

Yakovlev, A. Y., Tsodikov, A. D. and Asselain, B. (1996). *Stochastic models of tumor latency and their biostatistical applications*, World Scientific.

Youden, W. J. (1950). Index for rating diagnostic tests, *Cancer* **3**(1): 32–35.

Zheng, Y., Cai, T. and Feng, Z. (2006). Application of the time-dependent ROC curves for prognostic accuracy with multiple biomarkers, *Biometrics* **62**(1): 279–287.

Zheng, Y. and Heagerty, P. J. (2004). Semiparametric estimation of time-dependent ROC curves for longitudinal marker data, *Biostatistics* **5**(4): 615–632.

Zhou, X.-H. and Harezlak, J. (2002). Comparison of bandwidth selection methods for kernel smoothing of ROC curves, *Statistics in Medicine* **21**(14): 2045–2055.

Zhou, X.-H., Obuchowski, N. A. and McClish, D. K. (2011). *Statistical methods in diagnostic medicine*, John Wiley & Sons.

APPENDIX A

The cenROC Package

In this section, we introduce a user-friendly R package `cenROC` that makes use of the proposed time-dependent ROC curves and related summary measures estimation methods presented in this thesis readily available for researchers and practitioners. This package mainly contains six functions: two for ROC estimation, three for bandwidth selection, and the other for Youden index estimation. The arguments of each of these functions is detailed in the documentation of the package. A quick introduction of these functions, together with their usage, are presented below.

The `cenROC` function

The function `cenROC` is based on the methods presented in Chapter 3 and it estimates the time-dependent ROC curve and corresponding area under the ROC curve(AUC) for right censored time-to-event data. This function generates both non-smoothed (empirical) and smoothed ROC curve. The structure of parameters required in this function is

```
> library(cenROC)
Loading required package: condSURV
> args(cenROC)
function (Y, M, censor, t, U = NULL, h = NULL, bw = "NR",
method = "tra", ktype = "normal", ktype1 = "normal",
B = 0, alpha = 0.05, plot = TRUE)
NULL
```

The usage of this function is illustrated as follows:

```
> data(mayo)
> set.seed(1234)
> res=cenROC(Y=mayo$time, M=mayo$mayoscore5,
 censor=mayo$censor, t=365*6, B=1000)
> str(res)
List of 6
$ AUC:'data.frame':     1 obs. of   4 variables:
..$ AUC: num 0.877
..$ sd : num 0.0256
..$ LCL: num 0.825
..$ UCL: num 0.924
$ ROC:'data.frame':     151 obs. of   3 variables:
..$ ROC: num [1:151] 0.0593 0.1956 0.2789 0.3529 0.4176 ...
..$ LCL: num [1:151] 0.0353 0.1073 0.1595 0.2027 0.2453 ...
..$ UCL: num [1:151] 0.16 0.499 0.561 0.606 0.648 ...
$ U   : num [1:151] 0 0.00667 0.01333 0.02 0.02667 ...
$ Dt : num [1:312] 0.182 0.105 0.561 0 1 ...
$ M   : num [1:312] 11.25 10.14 10.1 10.19 9.77 ...
$ bw : num 0.25
```

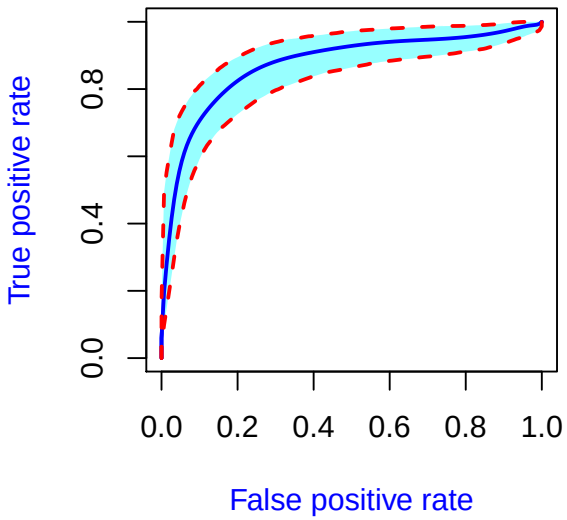

Figure A.1: Estimated time-dependent ROC curve with 95% confidence band obtained using percentile bootstrap approach.

The IntROC function

The function IntROC computes the time-dependent ROC curve and the corresponding time-dependent AUC for interval censored survival data. The ROC curves can be either empirical (non-smoothed) or smoothed with/without boundary correction. This function is based on the methods presented in Chapter 4. The structure and usage of this function are:

```
> args(IntROC)
function (L, R, M, t, U = NULL, method = "emp", method2 =
"pa", dist = "weibull", bw = NULL, ktype = "normal",
len = 151,  B = 0, alpha = 0.05, plot = "TRUE")
NUL
> data(hds)
> res <- IntROC(L=hds$L, R=hds$R, M=hds$M, t=2, len=500)
> str(res)
List of 6
$ AUC: num 0.752
$ ROC: num [1:500] 0 0 0 0 0 ...
$ U   : num [1:500] 0 0.002 0.00401 0.00601 0.00802 ...
$ Dt : num [1:238] 1 1 1 1 1 1 1 1 1 1 ...
$ M   : num [1:238] 7.99 9.28 8.79 8.33 6.77 ...
$ bw : logi NA
> res$AUC
[1] 0.7518023
```

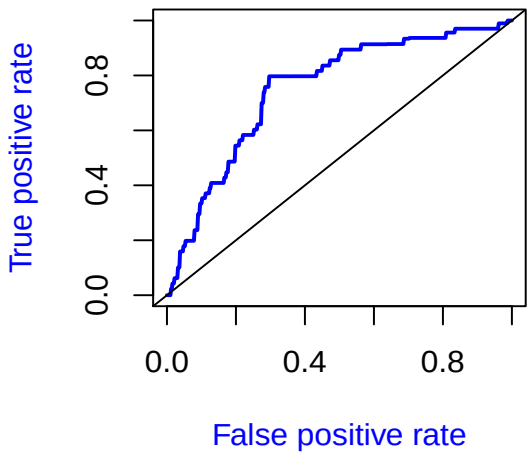

Figure A.2: Estimated time-dependent ROC curve for HDS data.

The youden function

The function **youden** is useful to compute the Youden index and the corresponding optimal cutoff point.

```
> args(youden)
function (est, plot = "FALSE")
NULL
```

Illustration

```
> res <- cenROC(Y=mayo$time, M=mayo$mayoscore5,
censor=mayo$censor, t=365*6)
> you <- youden(est=res, plot="TRUE")
> str(you)
'data.frame':   1 obs. of   4 variables:
$ Youden_index: num 0.627
$ cutopt         : num 6.27
$ sens           : num 0.8
$ spec           : num 0.827
```

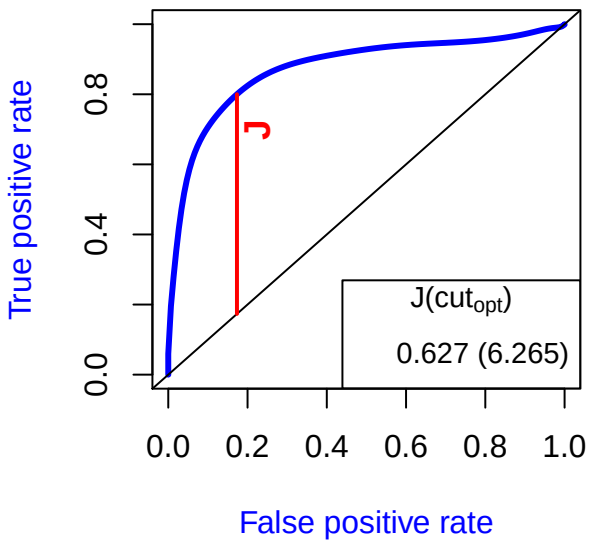

Figure A.3: Estimated time-dependent ROC curve with the Youden index and the corresponding optimal cutoff point.

The NR function

The function NR implements the normal reference bandwidth selection method given in equation (3.10) needed for smoothing the time-dependent ROC curve.

```
> args(NR)
function (X, wt, ktype = "normal")
NULL
```

Illustration

```
> set.seed(1234)
> X <- rnorm(100)
> wt <- runif(100)
> NR(X = X, wt = wt)$bw
[1] 0.3567449
```

The PI function

The function PI implements the plug-in bandwidth parameter selection method (see Section 3.3) needed for smoothing the time-dependent ROC curve.

```
> args(PI)
function (X, wt, ktype = "normal")
NULL
```

Illustration

```
> set.seed(1234)
> X <- rnorm(100)
> wt <- runif(100)
> PI(X = X, wt = wt)$bw
[1] 0.3001861
```

The CV function

The function CV calculates the bandwidth parameter using the cross-validation method given in equation (3.12) needed for smoothing the time-dependent ROC curve.

```
> args(CV)
function (X, wt, ktype = "normal")
NULL
```

Illustration

```
> set.seed(1234)
> X <- rnorm(100)
> wt <- runif(100)
> CV(X = X, wt = wt)$bw
[1] 0.3211515
```