Flexible Accelerated Failure Time modeling of multivariable time-to-event data

Menglan Pang

Doctor of Philosophy

Department of Epidemiology, Biostatistics and Occupational Health

McGill University, Montreal

August 11, 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

ABSTRACT

In survival analysis, the accelerated failure time (AFT) model provides an important alternative framework to the well-known Cox proportional hazards (PH) model. However, it is often difficult in real-life studies to correctly specify an appropriate distribution for the event time, which is required in a parametric AFT model. Moreover, accurate estimations of the covariate effects and survival probabilities require that the assumptions underlying a chosen regression model are consistent with the way each covariate affects survival. Methods for relaxing the PH and linearity assumptions in the Cox model have been well-studied. In contrast, there is sparse literature on addressing possible violation of the constant-over-time time ratios, and linearity of the covariate effects assumptions in the AFT model.

Therefore, the overall objectives of this thesis are (1) to enhance the conventional AFT model by avoiding the need to pre-specify the parametric distribution of the event time; (2) to relax the constant time ratio and linearity assumptions, regarding the covariate effects in the AFT model proposed in (1), and (3) to assess the potential improvements in the accuracy of the estimated survival probabilities, conditional on baseline prognostic factors, resulting from the developments proposed in point (2) above. This thesis aims to achieve these objectives in three corresponding manuscripts.

The first manuscript focuses on the development of a flexible spline-based AFT model that avoids the specification of the parametric family of the event time distribution. Spline smoothing techniques are used to model the baseline hazard function, allowing for the estimation of a variety of smooth and flexible shapes. Similarly, a recent semiparametric AFT model

developed by Komárek et al. also does not require such specification. Using comprehensive simulations, I evaluate the performance of the proposed method and compare it with the results from both (i) parametric AFT models and (ii) the alternative flexible approach proposed by Komárek et al. It is shown that both the proposed spline-based and the Komárek et al.'s AFT models provide accurate estimates of covariate effects, baseline hazards, and survival curves, while the proposed method yield more stable estimates of the hazard function. As expected, the mis-specified parametric AFT models result in biased estimates. The application of the proposed flexible AFT model is illustrated in a study of mortality in colon cancer.

In the second manuscript, a further flexible extension of the AFT model is developed to simultaneously model (i) the baseline hazard function of arbitrary shape, (ii) the non-linear effects (NL) of continuous covariates on the logarithm of the survival time, and (iii) the time-dependent (TD) time-ratios, that relax the AFT assumption regarding the covariate effects. Iterative alternating conditional (full) maximum likelihood estimation (MLE) algorithm is adapted to estimate the parameters in the flexible AFT model. In simulations, the proposed estimators are shown to accurately recover various plausible shapes of both the NL and TD curves, in multivariable settings. In addition, the survival curves, conditional on covariate vectors, can be accurately estimated, even in the presence of complex relationships between the covariates and the hazard. Furthermore, I apply the proposed method to re-assess the effects of prognostic factors on mortality after septic shock. Some continuous covariates are shown to have important NL and/or TD effects, which illustrate the additional insights offered by the flexible AFT model in real-life clinical studies.

In the third manuscript, both the new models developed in the first two manuscripts, and existing 'conventional' survival analytical models, are used to study the effects of baseline prognostic factors on survival in non-small cell lung cancer (NSCLC) patients, and to predict 1-year risk of death. To reduce model overfitting, backward selection based on the Akaike information criterion (AIC) is used to identify potential NL and/or TD effects in the flexible multivariable AFT model. Several other survival models, including the parametric Weibull

AFT model, the conventional Cox PH model and a flexible extension of the Cox model, are fit to the same data. Alternative models are then compared based on both goodness-of-fit and the prediction performance criteria. This study illustrates analytical challenges encountered when deciding about modeling strategy in complex, multivariable real-life survival analyses and shows the potential practical usefulness of the proposed flexible extensions of the AFT model.

RÉSUMÉ

En analyse de survie, les modèles de vie accélérée (AFT) représentent une alternative au modèle des risques proportionnels (PH) de Cox. Cependant, dans la pratique, il est souvent difficile de spécifier correctement la distribution du temps d'évènement, requise dans ces modèles AFT paramétriques. De plus, l'estimation avec précision de l'effet des covariables et de la probabilité de survie nécessite que les hypothèses sous-jacentes au modèle choisi soient en accord avec la manière dont chaque covariable affecte la survie. Beaucoup de développements ont été effectués permettant d'assouplir les hypothèses des risques proportionnels (PH) et de linéarité des effets du modèle de Cox. À l'inverse, très peu d'études portant sur la non-constance au cours du temps du rapport des temps ou sur la non-linéarité des effets des covariables, dans le cadre des modèles de vie accélérée, sont disponibles dans la littérature.

Par conséquent, les objectifs de cette thèse sont (1) d'améliorer le modèle AFT conventionnel en enlevant la nécessité de spécifier la distribution paramétrique du temps d'évènement; (2) d'assouplir l'hypothèse du rapport constant des temps et l'hypothèse de linéarité du modèle AFT proposé dans la premier objectif, et (3) d'évaluer les potentielles améliorations dans la précision de l'estimation de la probabilité de survie à partir des développements effectués dans le point (2) mentionné ci-dessus. Cette thèse a pour but d'accomplir ces trois objectifs en faisant l'objet de trois manuscrits, chacun référant à un objectif en particulier.

Le premier manuscrit se concentre sur le développement d'un modèle AFT flexible reposant sur l'utilisation des splines. Les techniques de lissage par fonctions splines sont utilisées pour modéliser le risque instantané de base, permettant l'estimation d'une variété de formes lisses

et flexibles. Un modèle AFT semi-paramétrique a récemment été développé par Komárek et al. À l'aide d'une étude de simulations, la performance de la méthode proposée est évaluée et comparée avec celle (i) des modèles AFT paramétriques et (ii) de l'approche alternative flexible proposée par Komárek et al. Il est montré que la méthode proposée ainsi que la méthode de Komárek et al fournissent des estimations non biaisées des effets des covariables, des risques instantanés de base et des courbes de survie. La méthode que nous proposons fournit cependant des estimations plus stables de la fonction du risque instantané. L'application de la méthode proposée est illustrée sur une étude de la mortalité suite à un cancer du côlon.

Dans le second manuscrit, une autre extension flexible du modèle AFT est développée afin de modéliser simultanément (i) le risque instantané de base, peu importe la forme, (ii) les effets non-linéaires des covariables continues, et (iii) les rapports des temps dépendant du temps (TD), ce qui permettrait d'assouplir les hypothèses sur les effets des covariables des modèles AFT conventionnels. L'algorithme itératif et alternatif de l'estimation du maximum de vraisemblance conditionnelle est adapté afin d'estimer les paramètres du modèle AFT flexible. À l'aide d'une étude de simulations, il a été montré que la méthode proposée permet l'estimation de façon précise d'une grande variété de formes pour les effets NL et TD, dans le cadre de modèles multivariés. De plus, les courbes de survie, conditionnelles aux covariables, sont estimées de façon précise, même en présence de relations complexes entre les covariables et le risque instantané. La méthode proposée a été appliquée afin de réévaluer les effets de facteurs pronostiques sur la mortalité après un choc septique.

Dans le troisième manuscrit, les modèles développés dans les deux précédents manuscrits sont utilisés pour étudier les effets des facteurs prognostiques au début du suivi sur la survie des patients atteints du cancer du poumon non à petites cellules, ainsi que pour prédire le risque de décès un an après le diagnostic. Pour réduire les problèmes de surajustement, la méthode de sélection descendante reposant sur le critère d'Akaike est utilisée pour identifier les potentiels effets NL et/ou TD dans le modèle AFT multivarié flexible. Plusieurs autres modèles de survie, incluant le modèle AFT paramétrique de Weibull, le modèle conventionnel

de Cox et une extension flexible du modèle de Cox, sont utilisés sur les mêmes données. Les différents modèles sont ensuite comparés en termes de qualité de l'ajustement mais aussi sur les critères de performance liés à la prédiction. Cette étude illustre les défis analytiques qui peuvent être rencontrés dans le choix des stratégies de modélisation dans le cadre d'analyses de survie complexes multivariées.

ACKNOWLEDGEMENTS

I would like to express my gratitude to many individuals who have supported and accompanied me during the journal of my PhD study. First and foremost, I would like to thank my supervisors, Drs. Michal Abrahamowicz and Robert Platt, who have provided me with countless guidance on my methodological research, precious collaborative opportunities, and generous financial support during the entire length of my training. I also appreciate their emotional support, professional advice, and their trust in me thorough out this process. This thesis could not have been possibly accomplished without them.

My appreciation also goes to my thesis committee member, Tibor Schuster, who has provided insightful comments on my research projects and invaluable help when I encounter computational problems. I am also grateful for the support of my colleagues at the Jewish General Hospital and McGill University Health Centre. Special thanks to Marie-Eve Beauchamp, Coraline Danieli, Chris Gravel, and Ehsan Karim for sharing your expertise in biostatistics and advice on career development. To my McGill Biostatistics peers, especially Gabrielle Simoneau, Janie Coulombe, Yi Lian, Yu Luo, Guanbo Wang, and Alex Levis, I was fortunate to be a part of this community and share this experience with such an amazing and talented group of people. I am also thankful for the efforts devoted by the faculty and administration of the department of Epidemiology, Biostatistics and Occupational Health to provide such a wonderful academic environment.

To the friends who have supported me throughout my PhD study especially during the most difficult times: Fei Wang, Siling Ou, Bo Chen, Debbie Dawson, and Wei Chen. Thank you for

being here with me whenever I need. Your continuous encouragement and friendship make this journey a memorable experience. I would also like to thank my accountability partners, Anqi Sun and Vivian Ugochi Ukah, for their support to help me stay motivated during my thesis writing till the end.

Finally, and most importantly, I would like to thank my family. To my mum and dad, words are certainly not enough to express my gratitude for your unconditional and endless love no matter where I go and no matter what decision I make. To my husband, Mianbo, thank you for your love, understanding and accompany. I am excited to open a new chapter of our life with you.

PREFACE

Statement of Originality

The work presented in this thesis constitutes an original contribution to the advancement of the statistical methodology in time-to-event analyses, focusing on the accelerated failure time (AFT) model. Although this method is relatively less frequently used in survival analysis comparing to the well-known Cox proportional hazards (PH) model, it merits further investigation given its utility for real-life studies, especially when the crucial assumptions of the Cox model are violated. The three manuscripts in this thesis address methodological gaps of the AFT model regarding its inherent assumptions and demonstrate the model applications and comparisons in real-life analyses.

Up to now, numerous semiparametric AFT models have been established to accurately estimate the covariate effects represented by the time ratio, while avoiding the parametric assumption about the event time or the error distribution. However, similar to the Cox PH model, most of the existing methods treat the baseline hazard function as a nuisance parameter, and thus the associated survival estimates are not available. The first manuscript addresses this limitation by developing a new semiparametric AFT model that uses flexible modeling for the baseline hazard. Full maximum likelihood with an alternating conditional estimation (ACE) algorithm is proposed to estimate the parameters that characterize the baseline hazard and the regression parameters for the covariate effects. Furthermore, using both simulation and real-life studies, I have systematically evaluated the performance of this new method and

compared it with the flexible smoothed error AFT model, an existing semiparametric approach that also facilitates the estimation of baseline hazard and survival functions.

In the second manuscript, I have further extended the spline-based AFT model developed in manuscript 1 to simultaneously relax the linearity and constant time ratio assumptions imposed by the AFT model. This challenge, to my knowledge, has not been addressed in the current literature, as these two assumptions are not widely recognized, and formulation and estimation of the model become increasingly difficult in complex multivariable settings. This developed comprehensive flexible AFT model is the first to accommodate the estimation of (i) non-linear (NL) effect of the continuous covariate, (ii) time-dependent (TD) time-ratio for all covariates, and (iii) individual survive curve, conditional on the estimated NL and TD effects, for specific covariate patterns. To validate this complex model, I have designed novel simulation studies with different scenarios that involve various plausible shapes of both the NL and TD functions. To reduce model overfitting, a data-adaptive procedure is proposed to select the relevant NL and TD effects.

Given several alternative models in time-to-event analysis, little guidance is available for analysts to decide on a model building strategy in practice. Real-life studies are included in each of the three manuscripts to demonstrate the applications of the developed methods. In the third manuscript, I compare the results of the new AFT methods developed in the first two manuscripts with the flexible Cox and other conventional survival models in a cancer prognostic study, where unknown but potentially complicate baseline hazards as well as NL and TD effects likely occur. This study illustrates the process of building flexible multivariable models in real-life studies, when different modeling approaches are being considered.

In summary, this research enhances the methodological development in the AFT models by incorporating estimations of complex covariate effects and baseline hazard and illustrates the importance of considering multiple analysis strategies when prior substantive knowledge is not available.

Contribution of Authors

Manuscript 1

Menglan Pang, Robert W. Platt, Tibor Schuster, and Michal Abrahamowicz. Spline-based Accelerated Failure Time Model. Submitted to *Statistics in Medicine*.

Manuscript 2

Menglan Pang, Robert W. Platt, Tibor Schuster, and Michal Abrahamowicz. Flexible Modeling of Time-dependent and Non-linear effects in Accelerated Failure Time Model. Submitted to *Biometrics*.

Manuscript 3

Menglan Pang, Robert W. Platt, Bruno Gagnon, and Michal Abrahamowicz. Comparison of Alternative Statistical Models for Re-assessing Survival in Advanced Non-small Cell Lung Cancer. In preparation for the *International Journal of Epidemiology*.

I was responsible for developing the details of the statistical methods, carrying out comprehensive simulation studies, writing the statistical programs to implement the new methods, conducting the real-life analyses, interpreting the results, and drafting the manuscripts. I was also responsible for conducting the literature review and writing the thesis. My doctoral work was performed under the guidance of my supervisor and thesis committee members. They were closely involved in defining the research topic and objectives of my thesis and provided valuable advice and feedback on the method development, simulation designs, analysis strategies, results interpretation, and drafts of each of the manuscripts.

Dr. Michal Abrahamowicz is a James McGill Professor in the Departments of Epidemiology, Biostatistics, and Occupational Health. Dr. Abrahamowicz is also a co-founder and co-Chair of the international STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative for improving the analyses of observational studies. He has many years of expertise in the development of novel flexible models for survival analysis and their applications in clinical and biomedical data. Dr. Michal Abrahamowicz provided both general direction and detailed guidance on methods proposal, new algorithm implementation, simula-

tion designs, and interpretation of complex results during all stages of my thesis work. He also provided extensive comments and revision on all the three manuscripts and the other chapters of the thesis.

Dr. Robert Platt is a Professor in the Departments of Pediatrics and Epidemiology, Biostatistics, and Occupational Health at McGill University. He is also the Albert Boehringer Chair in Pharmacoepidemiology and the Co-Principal Investigator of the Canadian Network for Observational Drug Effect Studies. Dr. Platt is also involved in all stages of my doctoral work. He provided guidance and feedback in the research proposal development, analytical strategies, methods applications, as well as presentation of study results. He also advised on the evaluation of the survival prediction performance and provided general comments on all the three manuscripts and the thesis.

Dr. Tibor Schuster, a member of my PhD thesis committee, is an Assistant Professor in the Department of Family Medicine at McGill University. He is also the Canada Research Chair in Biostatistical Methods for Primary Health Care Research. Dr. Schuster has expertise in computation-intensive statistical and machine learning methods. As such, he provided close guidance in the development and improvement of the estimation algorithms proposed in the first and second manuscripts and provided general comments on both manuscripts.

Dr. Bruno Gagnon is an Associate Professor in the Department of Family Medicine and Emergency Medicine at Université Laval. He is also a physician specialized in palliative care for patients with advance cancer. For the third manuscript, he provided clinical expertise in refining the research objectives and substantive knowledge of the prognostic factors for mortality in non-small cell lung cancer.

Statement of Financial Support

From the second to the fourth year of my doctoral studies, I was funded by a doctoral training award from the Fonds de Recherche du Québec-Santé (FRQS). I was also supported by stipends from my supervisors' Natural Sciences and Engineering Research Council of Canada (NSERC) grants for the last two years of my PhD training. For the fifth year, I

received a studentship award from the Research Institute of the McGill University Health Centre (RI-MUHC). I was also partially supported by funding from the Canadian Network for Observational Drug Effect Studies (CNODES) throughout my PhD studies. I received McGill MedStar award in recognition of excellent research, and Canadian Statistical Sciences Institute (CANSSI) travel awards, Departmental Graduate Research Enhancement and Travel (GREAT) awards, International Conference on Pharmacoepidemiology (ICPE) scholarship to present my doctoral work at conferences.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACE	Alternating Conditional Estimation
AFT	Accelerated Failure Time
AIC	Akaike Information Criterion
BIC	Bayesian Information Criteria
CI	Confidence Interval
df	Degrees of Freedom
EHR	Extended Hazard Regression
HR	Hazard Ratio
HARE	Hazard Regression
LRT	Likelihood Ratio Test
MLE	Maximum Likelihood Estimation
NL	Non Linear
NSCLC	Non-small Cell Lung Cancer
PH	Proportional Hazards
PO	Proportional Odds
rMSE	root Mean Squared Error
SD	Standard Deviation
TD	Time Dependent

CHAPTER 1

Introduction

In epidemiological and clinical analyses of time-to-event data, the Cox Proportional Hazards (PH) model allows an efficient way to estimate the hazard ratio's (HR) for the covariate effects without the need to specify the baseline hazard function [1]. Because in many real-life studies, the underlying baseline hazard function is unknown, this probably is one of the main reasons for the predominant use of the Cox PH model in real-life survival analyses. However, an unintended side effect of the use of the Cox PH model is a lack of emphasis on the hazard function, which can be informative [2, 3]. Moreover, the estimation of the hazard function could further facilitates the estimation of individual-specific survival curve, which provides essential information for physicians and patients about prognosis and treatment options [4, 5].

The accelerated failure time (AFT) model was proposed as an alternative to the PH model for survival analysis [6]. In contrast to the PH model, where the HR represents the covariate effect on the instantaneous risk, the AFT model postulates a direct relationship between the covariate values and the event times [7]. It is similar to the classical linear regression in that it implies modeling the natural logarithm of the event time as the response variable. However, in contrast to linear regression, the AFT model accounts for censoring of time-to-event outcomes. The regression parameter in the AFT model is the time ratio, which is a multiplicative factor that quantifies the acceleration or deceleration of the event time associated with a given change in the covariate value. If the model is correctly specified, the estimation of the time ratio can

offer a clinically meaningful interpretation of the effects of treatment and prognostic factors [6, 8, 9].

In contrast to the Cox PH model, parametric AFT model requires that the event time distribution to be specified and provides estimation of the hazard and survival function according to the parametric assumption [7]. Weibull, exponential, log-normal, log-logistic, and gamma distributions are common choices. However, misspecification may lead to biased estimates of covariate effects and survival [10, 11]. Additionally, in most complex real-life studies, substantive knowledge is often insufficient to identify an appropriate event time distribution, and the fact that observed survival is affected by several covariates makes it even more difficult to assess the baseline hazard, corresponding to a specific (reference) covariate vector.

Several different semiparametric approaches have been proposed for the estimation of the covariate effects in the AFT model while avoiding the need to pre-specify the event time distribution. Classical semiparametric approaches include least squares regression-based [12, 13, 14], and linear-rank-test-based estimators [15, 16, 17]. However, these semiparametric estimating equations have potentially multiple solutions, leading to computational problems, especially with a large number of covariates [6, 18, 19]. Other semiparametric AFT approaches include a weighted least-absolute-deviations method [20], and a semiparametric AFT mixture cure model [21, 22, 23]. However, none of these aforementioned methods facilitates the estimation of the hazard function and the survival curves.

On the other hand, among the few other methods that did allow for these estimations, Etezadi-Amoli and Ciampi have proposed a more general extended hazard regression (EHR) [24, 25] that includes both the PH and AFT models as special cases. However, to ensure that the baseline hazard is nonnegative, complex constrained optimization is required for parameter estimation, leading to inefficiency of the estimation algorithm [25]. Furthermore, no statistical software is available to implement their method [25]. Komárek et al. have developed a semiparametric AFT model for arbitrarily censored data [26]. A mixture of numerous Gaussian densities is used to approximate the error distribution, and thus a large number of

parameters are involved in the estimation procedure of this method. To control the smoothness of the fitted error distribution and to reduce model overfitting, a penalty term is applied to the log-likelihood and is selected based on cross-validation [26]. Simulation studies, under two settings with respect to the true error distribution, have shown satisfactory performance regarding the accuracy of the covariate effect estimates and the fitted error distribution [26]. It would be useful to extend the simulation studies in more scenarios considering a variety of different data generating mechanisms and, at the same time, evaluate the resulting hazard and survival estimation.

Although the aforementioned semiparametric AFT models attempt to relax the parametric assumption about the event time distribution, they rely on two other important implicit assumptions. First, similar to the PH assumption imposed in the Cox PH model, the time ratio in the AFT model is also restricted to be constant across the follow-up (constant time ratio assumption) [3]. Second, the formulation of the conventional AFT model also assumes that the continuous covariates affect the logarithm of the event time linearly (linearity assumption) [27]. It is essential to test these assumptions and account for their possible violations in epidemiological and clinical studies. Indeed, misspecification of the functional form and violation of the PH assumption in the Cox model can yield inaccurate estimates and invalid inference [28, 29, 30]. Moreover, important non-linear (NL) effects and time-dependent (TD) HRs have been already revealed for different prognostic factors and treatments, in many real-life studies of mortality of coronary heart disease [31, 32], colon cancer [33], non-small cell lung cancer (NSCLC) [34, 35], primary biliary cirrhosis [36], or prognosis of breast cancer [37, 38].

Several extensions of the Cox model have been developed to relax the PH and/or the linearity assumptions. For example, Sauerbrei et al. have carried out a multivariable fractional polynomial procedure [39], and Abrahamowicz and MacKenzie have proposed a regression B-spline approach [40] to select and simultaneously model the TD and NL effects. Furthermore, Wynant and Abrahamowicz have developed an extended flexible Cox model for the full maximum likelihood estimation (MLE) of individual survival curve conditional on the potential

NL and TD effects [41]. Since the linearity and the constant time ratio assumptions imposed by the conventional AFT model are as arbitrary as the linearity and constant HR assumptions imposed in the Cox PH model, it is necessary to check the corresponding assumptions and account for possible NL and TD effects in the AFT framework.

AFT partial linear models, based on spline smoothing, have been proposed to relax the linearity assumption. In particular, Orbe et al. [27] have adopted natural splines, Zou et al. [42] have used penalized splines, and Xue et al. [43] piecewise linear function to approximate the NL functional form of the continuous covariate. No specification of the event time distribution is needed; however, all these methods rely on *a priori* (implicit) constant time ratio AFT assumption. Indeed, to the best of my knowledge, the extended linear hazard regression proposed by Elsayed et al. [44] is the only method that can incorporate potentially time-dependent time ratios, by including an interaction between the covariate and follow-up time in the regression model. However, their approach seems to impose linear changes in log hazard ratio, whereas more flexible shapes of the TD effects may be often required, based on evidence from real-life studies [45, 46]. Moreover, Elsayed et al. do not discuss how the hazard and survival estimates, conditional on covariates, can be estimated [44].

In conclusion, whereas the AFT model starts to receive increasing attention in statistical literature, several methodological challenges remain to be addressed in order to allow accurate AFT analyses of complex multivariable time-to-event data encountered in most real-life applications. In particular, it is important to develop flexible extensions of the multivariable AFT model to (i) relax the parametric assumptions about the event time distribution, (ii) allow for the modeling of both the TD and the NL effects of continuous variables, and (iii) provide accurate individual hazard and survival estimation conditional on relevant, potentially NL and/or TD, covariate effects.

Furthermore, in real-life time-to-event analyses, it is challenging to decide which of the alternative regression models is most consistent with the empirical data. For example, except when event times follow a Weibull or exponential distribution, the Cox's PH and AFT constant

time ratio assumptions cannot be both correct, and it is possible that neither of them is fully satisfied. However, it is far from clear what empirical criteria can be used to provide a reliable comparison of the AFT vs. PH models, or their flexible extensions. Furthermore, there is little insights or guidance regarding the model building strategy in the context of AFT modeling of complex multivariable survival data, possibly including NL and/or TD covariate effects. Finally, there is little empirical evidence on head-to-head comparisons of alternative models for survival analysis.

CHAPTER 2

Literature Review

2.1 Survival analysis: general formulation

In epidemiological and medical research, the time to a specific event is commonly defined as the outcome of interest, such as the time from heart attack to death, or time from diagnosis of cancer to relapse [47, 48, 49, 50]. Let T denote the random variable for the time to event; the distribution of T can be specified in many intercorrelated ways. The survival function, hazard function, and the probability density functions are particularly useful in survival applications [7]. The survival function, which is defined as the probability that the event occurs later than some specific t, is given by:

$$S(t) = P(T > t) = \int_t^{\infty} f(u)du \tag{2.1.1}$$

where $f(\cdot)$ is the probability density function of T [7]. The survival function is a non-increasing function of t and $S(0) = 1$ assuming everyone is alive at time zero where the follow-up begins, and the event does not occur immediately. $S(t) = 0$ as $t \to \infty$, assuming the event would occur eventually if the follow-up time is long enough. The cumulative density function is given by $F(t) = P(T \leq t) = 1 - S(t)$. The instantaneous risk of the event at any given point in time is defined by the hazard function, denoted $\lambda(t)$, which describes the event rate at time

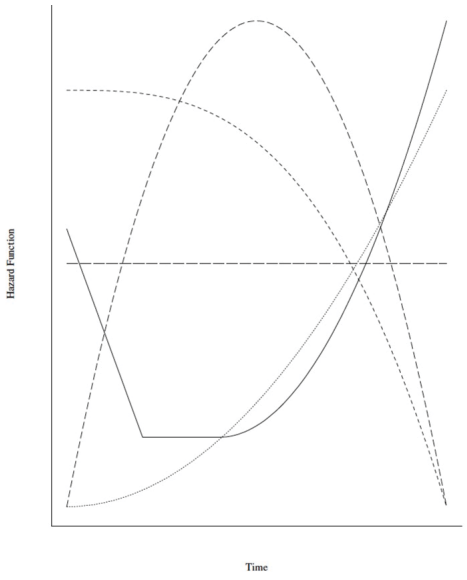
t conditional on event-free survival up to t:

$$\lambda(t) = \lim_{\triangle t \to 0} \frac{P(t \leq T < t + \triangle t | T \geq t)}{\triangle t} = \frac{f(t)}{S(t)}. \tag{2.1.2}$$

The cumulative hazard function, denoted by $\Lambda(t)$, represents the total amount of risk that has been accumulated up to time t [51], and is defined as:

$$\Lambda(t) = \int_0^t \lambda(u) du. \tag{2.1.3}$$

The hazard function is always non-negative, $\lambda(t) \geq 0$ for any $t \geq 0$, and $\Lambda(t) = \infty$ as $t \to \infty$. The hazard function can take various shapes that reflect different the underlying mechanism of the occurrence of the event and describe how the chance of experiencing the event changes with time. Several generic shapes of hazard function are depicted in Figure 2.4 (page 28) of Klein and Moeschberger [52]. For instance, it can be a non-monotonic function when there is an increased risk of death soon after the surgery of removing a tumor, but once patients survive up to a certain time point, the risk of death is decreasing afterward.



Shapes of hazard functions. Constant hazard (——); increasing hazard (------); decreasing hazard (- - - - -); bathtub shaped (——); humpshaped (——).

Figure 2-1: Shapes of hazard functions

There is a one-to-one relationship linking the hazard function with the probability density function and the survival function, as shown in (2.1.2). It also follows that:

$$\lambda(t) = \frac{F'(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d\log(S(t))}{dt}, \tag{2.1.4}$$

and

$$S(t) = \exp\left(-\Lambda(t)\right). \tag{2.1.5}$$

Censoring occurs when the subject's event is not observable, and the event time cannot be accurately measured. There are three types of censoring in time-to-event data, i.e., (i) left censoring, when the event is not observed but is known to occur before a certain time point, (ii) interval censoring, when the event is known to occur between a time interval, and (iii) right censoring, when the event is unobserved but potentially occurs after a certain time point [7]. Right censoring is the most common type of censoring in healthcare research and can occur for several reasons, such as loss to follow-up, termination of the study before everyone has the event or competing risks where the subject can experience the event caused by other reason rather than the one under study. For example, if death from lung cancer is the outcome of interest, dying from a stroke is a competing event that may be considered as censoring. This thesis focuses on developing new survival models for data subject to right censoring.

In the presence of right-censoring, we observe the survival time as a pair of data t_i, δ_i for subject i, where $t_i = \min(T_i, C_i)$ is the observed time that occurs first between the event time and the censoring time and $\delta_i = I(T_i \leq C_i)$ is the indicator whether the event has occurred or not before censoring. The likelihood of the given observed data can be constructed using the density function and the survival function. For the individual who had the event, the observation contributes to the likelihood as the density of the event at t_i, i.e., $P(T = t_i) = f(t_i)$; while for the individual who was censored at t_i, the only information we have is that the event time exceeds t_i; therefore the observation contributes to the likelihood as the survival function at t_i, i.e., $P(T > t_i) = S(t_i)$ [3, 7]. For a random sample of n subjects who are assumed to be

independent and identically distributed, the likelihood is expressed by:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i; \boldsymbol{\theta}, \boldsymbol{X}_i)^{\delta_i} S(t_i; \boldsymbol{\theta}, \boldsymbol{X}_i)^{1-\delta_i} = \lambda(t_i; \boldsymbol{\theta}, \boldsymbol{X}_i)^{\delta_i} S(t_i; \boldsymbol{\theta}, \boldsymbol{X}_i) \tag{2.1.6}$$

where $\boldsymbol{\theta}$ is the vector of the parameters in the density function, and $\boldsymbol{X}_i$ is a vector of observed covariates associated with the i^{th} subject. Another assumption we make throughout this thesis is independent censoring, which means that conditional on the covariates being included in the model, the distribution of the unobserved event time among those who are censored is the same as those who remained in the study [7, 53].

The Kaplan-Meier curve is a non-parametric estimator for estimating the survival function for the time-to-event data. The Kaplan-Meier estimator, also called the product-limit estimator, is given by [54]:

$$\hat{S}_{KM}(t) = \prod_{i:t_i \leq t} (1 - \frac{d_i}{n_i}) \tag{2.1.7}$$

where t_i is the i^{th} distinct uncensored event time, d_i is the number of events occur at t_i, and n_i is the number of subjects that are still at risk of experiencing the event (subjects who have not yet had the event or been censored) at t_i. The cumulative hazard, using the same notation, can be estimated by the Nelson-Aalen estimator [54],

$$\hat{\Lambda}_{NA}(t) = \prod_{i:t_i \leq t} \frac{d_i}{n_i} = \prod_{i:t_i \leq t} \hat{\lambda}_i \tag{2.1.8}$$

which is a non-parametric step function with increments being the empirical hazard estimate of $\hat{\lambda}$. The survival estimate based on the Nelson-Aalen estimator of the cumulative hazard function is an alternative non-parametric estimator of the survival function:

$$\hat{S}_{NA}(t) = \exp\big(-\hat{\Lambda}_{NA}(t)\big). \tag{2.1.9}$$

Note that both the Kaplan-Meier estimator and the Nelson-Aalen estimator are discrete right-continuous step functions that jump at the observed event times. Both estimators of the survival function or the cumulative hazard function are consistent and asymptotically equivalent [52].

Before introducing the accelerated failure time (AFT) model, I will first review the Cox Proportional Hazards (PH) model, which is the most commonly used model for survival data in biomedical research.

2.2 The conventional Cox proportional hazards (PH) model

The Cox PH model [1] is the de facto default method to analyze time-to-event data in medical research. This model allows us to estimate and make inferences about the effect of covariates without assuming any form for the baseline hazard function or the distribution of the event time. The Cox PH model is specified as:

$$\lambda(t|\boldsymbol{X}) = \lambda_0(t) \exp\left(\boldsymbol{\beta}^{*T}\boldsymbol{X}\right) \tag{2.2.1}$$

where $\lambda(t|\boldsymbol{X})$ is the hazard function at time t given the covariate vector $\boldsymbol{X}$, $\lambda_0(t)$ is the baseline hazard function when $\boldsymbol{X} = \boldsymbol{0}$ and $\boldsymbol{\beta}^*$ is a vector of regression coefficients associated the covariate vector $\boldsymbol{X}$, representing the logarithm of hazard ratio (HR) for one unit increase in covariate x while holding the other covariates constant. It follows that the survival function conditional on covariate $\boldsymbol{X}$ can be expressed as:

$$S(t|\boldsymbol{X}) = \exp\Big[-\int_0^t \lambda_0(u) \exp\left(\boldsymbol{\beta}^{*T}\boldsymbol{X}\right) du\Big] = S_0(t)^{\exp\left(\boldsymbol{\beta}^{*T}\boldsymbol{X}\right)} \tag{2.2.2}$$

where $S_0(t) = \exp\big[-\int_0^t \lambda_0(u)du\big]$.

2.2.1 Partial likelihood estimation for the Cox model

Let $t_1 \leq \cdots \leq t_k$ denote k uncensored event time (assuming no ties), and j denote the individual that has the event at t_j, the partial likelihood is given by [1]:

$$L(\boldsymbol{\beta}^*) = \prod_{j=1}^{k} L_j(\boldsymbol{\beta}^*) = \prod_{j=1}^{k} \frac{\lambda_0(t_j) \exp\left(\boldsymbol{\beta}^{*T}\boldsymbol{X}_j\right)}{\sum_{i \in R_{t_j}} \exp\left(\boldsymbol{\beta}^{*T}\boldsymbol{X}_i\right)} \tag{2.2.3}$$

where R_{t_j} denotes the risk set at t_j that consists of individuals who are still at risk of having the event at t_j. Maximizing the partial likelihood can provide estimation of $\boldsymbol{\beta}^*$. Since the baseline hazard $\lambda_0(t_j)$ cancels out in the numerator and the denominator at each time t_j, $j = 1,\cdots, k$, in the partial likelihood, no specification of the baseline hazard function is necessary to estimate $\boldsymbol{\beta}^*$ for the Cox model.

2.2.2 Estimation of the survival functions under the Cox model

The partial likelihood estimation provides an efficient way to estimate HR. However, it is not adequate to draw a conclusion about the effect of a covariate without knowing the absolute risk. On the other hand, the survival probabilities over time provide a different perspective on the event under study. Therefore, the survival function and HR can complement each other in delivering a better understanding of the substantive implication of the results [5, 55]. Because the baseline hazard function $\lambda_0(t_j)$ is left unspecified as a nuisance parameter, the Cox PH model does not provide a direct estimate of the baseline survival function $S_0(t)$ when all covariates are equal to 0 or the conditional survival function $S(t|\boldsymbol{X})$ for subjects with specific covariate value $\boldsymbol{X}$. Breslow in 1972 has proposed an estimator of the baseline cumulative hazard function $\Lambda_0(t)$ and the regression parameter $\boldsymbol{\beta}^*$ by maximizing the full likelihood function under the PH assumption [56, 57]:

$$L(\boldsymbol{\beta^*}, \Lambda_0) = \prod_{i=1}^{n} \left\{\lambda_0(t_i) \exp\left(\boldsymbol{\beta}^{*T}\boldsymbol{X}\right)\right\}^{\delta_i} \exp\left\{-\int_0^t \lambda_0(u) \exp\left(\boldsymbol{\beta}^{*T}\boldsymbol{X}\right) du\right\} \tag{2.2.4}$$

where $\lambda_0(t)$ was treated as a piecewise constant between uncensored event times. It has been shown that the maximum likelihood estimation (MLE) of $\boldsymbol{\beta}^*$ is equivalent to that from Cox's partial likelihood estimation, and the Breslow estimator of $\Lambda_0(t)$ is given by:

$$\hat{\Lambda}_0(t) = \sum_{i:t_i \le t} \frac{d_i}{\sum_{l \in R_{t_i}} \exp\left(\boldsymbol{\beta}^{*T}\boldsymbol{X}_l\right)}. \tag{2.2.5}$$

The corresponding estimator for the conditional survival function associated with the covariate vector $\boldsymbol{X}$ is:

$$\hat{S}(t|\boldsymbol{X}) = \exp\left\{ - \hat{\Lambda}_0(t) \exp\left(\boldsymbol{\beta}^{*T} \boldsymbol{X}\right)\right\}. \tag{2.2.6}$$

The Breslow estimators of $\Lambda_0(t)$ and $\hat{S}(t|\boldsymbol{X})$ have been implemented in most of the statistical packages, facilitating the estimation of survival functions in real-life applications. However, these functions, which contain intrinsic information about patient's diagnostic information, are rarely reported in studies published in the major general medical journals [5]. This may be because investigators tend to neglect other important characteristics in survival beyond the HR from the Cox model, where the baseline hazard is considered a nuisance parameter. Moreover, the resulting $\Lambda_0(t)$ and $S(t|\boldsymbol{X})$ are estimated as step functions that jump at each distinctive event time, and this may seem less clinically plausible than functions that are smooth in time.

2.2.3 The assumptions of the Cox PH model

The conventional Cox PH model imposes two assumptions. First, the effect of each continuous covariate on the logarithm of the hazard is assumed to be linear (linearity assumption). Martingale residuals and cumulative martingale residuals can help in investigating the functional forms of the continuous variables included in the model [52, 58]. However, this approach is ad hoc, and the linearity assumption is often taken for granted in many real-life applications. Secondly, the association between the covariates and the hazard of the event is measured by a hazard ratio (HR): $\exp\left(\boldsymbol{\beta}^{*T}(\boldsymbol{X}_i - \boldsymbol{X}_j)\right)$, comparing two vectors of covariate values, $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$. The hazard ratio does not depend on time. This constraint is referred to as the proportional hazards (PH) assumption of the Cox model, implying the effect of each covariate on survival does not change over time. Several graphical methods and statistical tests can be used to assess the PH assumption [59, 60, 61], such as the log-log curve, the Schoenfeld residual plot, Schoenfeld residual test, and tests based on interactions between time and the covariates. However, the graphical assessments rely on subjective decisions, whereas the statistical tests

depend on the choice of time dependence and maybe underpowered to detect PH violation with heavy censoring [62]. Moreover, in practice, this assumption is rarely tested in applications using the Cox model [63], although the violation of the PH assumption (constant hazard ratio over time) implies important limitations. In situations where the PH assumption does not hold, the Cox model-based HR estimate averages the effects across the follow-up duration and thus may lead to misleading conclusions [29]. On the other hand, time-specific HRs, estimated among those who survived until a given time, are subject to built-in selection bias [29]. This bias is due to the differential depletion of susceptible between exposure groups when a risk factor associated with the susceptibility is not accounted for in the Cox model. This problem can also be explained by using a causal diagram [64, 65]; in the presence of an unobserved risk factor, conditioning on survivors opens a non-causal pathway between the exposure and the risk factor, and thus induces a specific form of collider bias.

Another limitation of the Cox model concerns the difference between conditional and marginal HRs. When omitting a covariate, like the odds ratio, the HR is subject to non-collapsibility due to the non-linearity of the partial likelihood with respect to covariates [66, 67, 68, 69]. This leads to difficulty in a direct comparison between the conditional and marginal effects, even in the absence of confounding.

2.3 Overview of the accelerated failure time (AFT) model

The AFT model has been suggested as a useful alternative to the Cox PH model [6, 70]. In a typical AFT model, the natural logarithm of the event time, $\log T$, is modeled as a linear function of the covariate vector $\boldsymbol{X}$ [7]:

$$\log T = -\boldsymbol{\beta}^T \boldsymbol{X} + W \tag{2.3.1}$$

where W, independent of the covariate, is a random error term, and $\boldsymbol{\beta}$ is the vector of regression parameters. A positive value of β indicates that an increase in the covariate X would accelerate

or shorten the time to the event, corresponding to a harmful effect from a positive log HR obtained by the Cox model. On the other hand, a negative value of β implies a prolongation of the time to event. A slightly different parameterization can be found [7, 52],

$$\log T = \mu + \boldsymbol{\theta}^T \boldsymbol{X} + \sigma\varepsilon \tag{2.3.2}$$

where $\boldsymbol{\theta} = -\boldsymbol{\beta}$, μ is the intercept, σ is a scale parameter such that the error ε has a standard distribution when it belongs to a location-scale family. Although the parameterization (2.3.2) is implemented by most of the statistical packages (e.g., the 'survreg' function within the R 'survival' package) for fitting the parametric AFT model [71], modeling the additional parameter μ and σ does not provide additional information regarding the event time distribution or baseline hazard. Hence, we consider the random error term $W = \mu + \sigma\varepsilon$ without scaling in the AFT model throughout this thesis.

According to the one-to-one relationship between the probability density function and hazard function, we can derive the specification of the AFT model with respect to the specification of the hazard function. Exponentiating both sides of the equation (2.3.1) gives:

$$T = \exp\left(-\boldsymbol{\beta}^T \boldsymbol{X}\right) T_0 \tag{2.3.3}$$

where $T_0 = \exp(W)$ is the random variable that represents the baseline event time distribution with the probability density function $f_0(t)$, cumulative density function $F_0(t)$, survival function $S_0(t)$, and baseline hazard function $\lambda_0(t)$. Using a transformation, $T_0 = \exp\left(\boldsymbol{\beta}^T \boldsymbol{X}\right) T$, the density function of T conditional on $\boldsymbol{X}$ is given by:

$$f(t|\boldsymbol{X}) = f_0\big(\exp\left(\boldsymbol{\beta}^T \boldsymbol{X}\right) t\big) \exp\left(\boldsymbol{\beta}^T \boldsymbol{X}\right). \tag{2.3.4}$$

The corresponding cumulative density function and the survival function of T are:

$$F(t|\boldsymbol{X}) = \int_0^t f_0\big(\exp\left(\boldsymbol{\beta}^T \boldsymbol{X}\right) u\big) \exp\left(\boldsymbol{\beta}^T \boldsymbol{X}\right) du = F_0\big(\exp\left(\boldsymbol{\beta}^T \boldsymbol{X}\right) t\big), \tag{2.3.5}$$

and

$$S(t|\boldsymbol{X}) = 1 - F(t|\boldsymbol{X}) = S_0\big(\exp(\boldsymbol{\beta}^T\boldsymbol{X})t\big). \tag{2.3.6}$$

Therefore, it follows that the hazard function of the failure time T given the covariate vector $\boldsymbol{X}$ can be expressed by:

$$\lambda(t|\boldsymbol{X}) = \frac{f(t|\boldsymbol{X})}{S(t|\boldsymbol{X})} = \frac{f_0\big(\exp(\boldsymbol{\beta}^T\boldsymbol{X})t\big)\exp(\boldsymbol{\beta}^T\boldsymbol{X})}{S_0\big(\exp(\boldsymbol{\beta}^T\boldsymbol{X})t\big)}$$

$$= \exp(\boldsymbol{\beta}^T\boldsymbol{X})\lambda_0\big(\exp(\boldsymbol{\beta}^T\boldsymbol{X})t\big). \tag{2.3.7}$$

In the AFT model, for example, for a binary treatment variable, it is supposed that the time to event for a subject is accelerated or decelerated by the time ratio e^{β} comparing $X = 0$ vs. $X = 1$ (control vs. treated). Equivalently, the expressions equations (2.3.3) and (2.3.6) indicate that the median or any other quantiles of the survival time distribution in the treatment group is $e^{-\beta}$ times the equivalent quantile in the control group.

2.4 Comparison between the Cox PH and AFT model

The difference between the Cox PH model (2.2.1) and the AFT model (2.3.1) is immediately reflected by the way covariates affect the hazard function. The effect of a covariate is multiplicative on the hazard scale in the PH model, whereas it is multiplicative on the time scale in the AFT model. Figure 2-2 illustrates the comparison of the Cox PH and the AFT model for the effect of a binary treatment X on the hazard (left panel) scale and the log hazard (right panel) scale, respectively [7]. The gray curves represent the baseline hazard in the control group, while the red and blue curves show the hazards for the lower-risk treatment group implied by, respectively, the PH model and the AFT model. The treatment is assumed to have a protective effect on survival under both models with an HR and a time ratio both equal to $\frac{2}{3}$. This Figure shows clearly that under the PH model, the covariate decreases the baseline hazard by the constant log HR, as reflected by a vertical shift. In contrast, under the

AFT model, the covariate not only decreases the baseline risk by the log time ratio vertically but also acts on the time scale by shifting the hazard towards the right. It is important to note that for a given covariate, the PH model and AFT model cannot both be correct unless the event time distribution is Weibull or exponential, the only special cases where the PH model and AFT model coincide [7, 72]. In other words, the AFT model implies non-proportional hazards, whereas the PH model implies a non-constant time ratio. Nevertheless, because of the common effect in the vertical direction on the log hazard, no large disparity would be expected between the resulting HR and time ratio assuming that the same covariates with the same functional forms are included in both models.

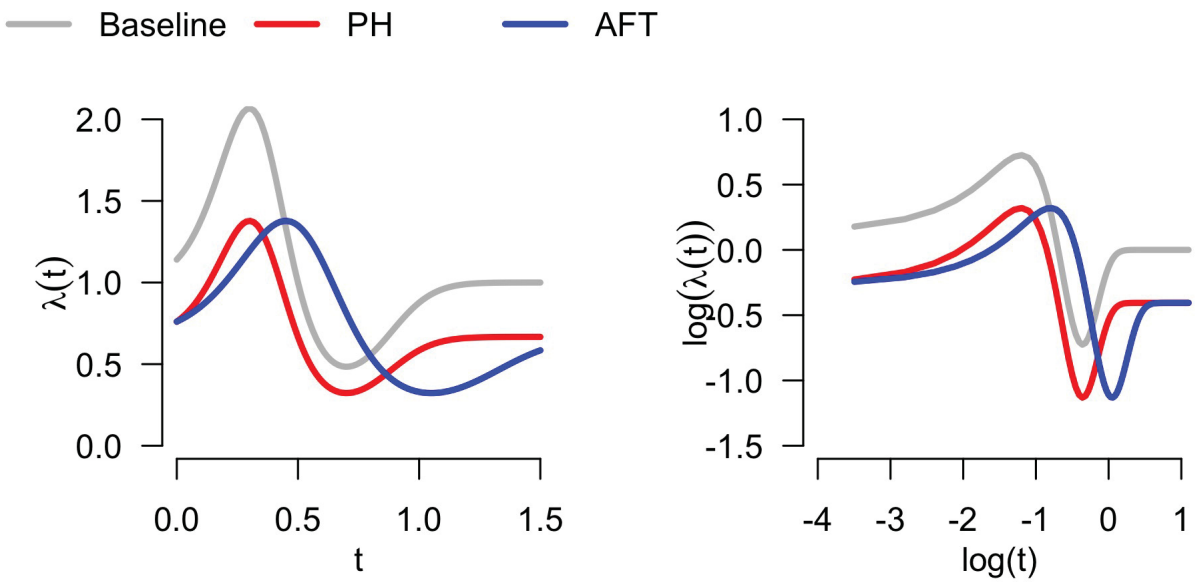


Figure 2-2: Comparison of PH and AFT model with respect to the hazard functions

Figure 2-3 shows the corresponding comparison with respect to the survival functions using the same example as in Figure 2-2. The difference between the PH and AFT models has important implications. If the data generating mechanism indeed follows the AFT model (the blue curves in Figure 2-2), where the HR is clearly not constant over time, the single estimated HR by the default Cox PH model would yield inaccurate effect estimates. Moreover, the survival probabilities are underestimated before $t = 0.5$ and are overestimated afterward.

Vice versa, the time ratio from the AFT model may provide a misleading conclusion if the covariate effect truly conforms to a PH model where a constant time ratio is unlikely. Therefore, one should be careful about the assumptions implied by both models in order to deliver valid prognostic information on survival. The violation of the constant time ratio assumption in the AFT model is explained in section 2.5.2. It has also been suggested that investigators should apply both the PH and AFT models and choose the most appropriate approach for the particular dataset under study [73].

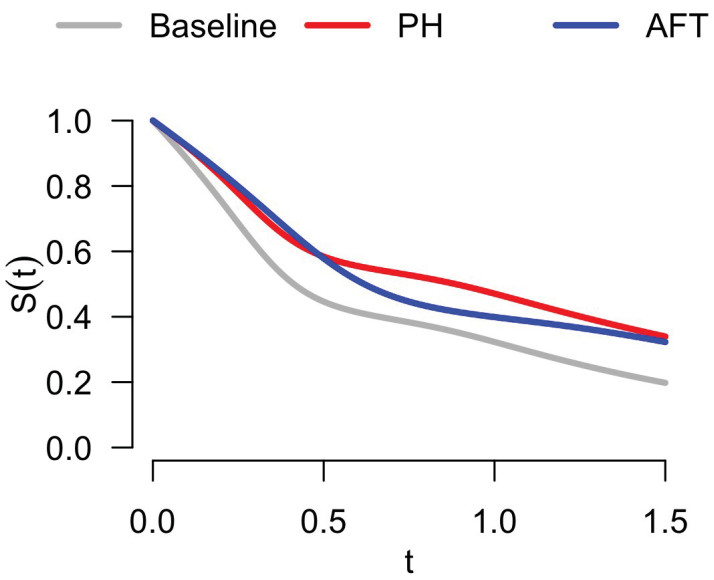


Figure 2-3: Comparison of PH and AFT model with respect to the survival functions

As previously noted in section 2.3, the effects of the covariates in the AFT model are interpreted based on the time scale. Due to the different mechanisms regarding how covariates alter the underlying hazard function, the HR and time ratio of the same covariate from the PH and AFT models when analyzing the same dataset are not generally comparable except for a Weibull or exponential baseline hazard. The time ratio resulting from the AFT model postulates a direct relationship between the covariates and event time and thus has arguably more intuitive interpretation compared to the HR from the Cox model [9, 73, 74]. Indeed, in his conversation with Nancy Reid in 1994, Sir David Cox remarked that "of course, another

issue is the physical or substantive basis for the proportional hazards model. I think that's one of its weaknesses, that accelerated life models are in many ways more appealing because of their quite direct physical interpretation, particularly in an engineering context" [75].

2.5 Assumptions in the AFT model

While there is growing literature regarding the assessment of the linearity and PH assumptions and strategies for relaxing these assumptions in the context of the Cox model, the discussion about the assumptions imposed by the AFT model is minimal. Similar to the Cox PH model, two assumptions are imposed by the conventional AFT model, namely the linearity assumption and the constant time ratio assumption.

2.5.1 The linearity assumption in the AFT model

The log-linear form of the AFT model (2.3.1) suggests that continuous covariates affect the logarithm of the event time linearly. This would imply that, for example, the time ratio comparing median survival times for a 10-year age difference is the same when comparing (a) subgroups that are 70-year old vs. 60-year-old, and (b) subgroups that are 30-year old vs. 20-year old. In most clinical and epidemiological studies, continuous variables are often categorized, or the linearity assumption is often accepted *a priori* [76, 77, 78]. While categorization may introduce a loss of efficiency and residual confounding [76, 79, 80], imposing an incorrect log-linear relationship may lead to important biases and misleading conclusions [34, 77, 81]. Indeed, the violation of the linearity assumption has been found for many prognostic and risk factors in several studies using the Cox model [31, 36, 37, 46, 82]. Moreover, it has been suggested that this violation of the effect of a given variable may induce biased estimates for other prognostic factors included in the model [83]. Thus, it is important to check the linearity assumption in the AFT model and develop a method to allow for non-linear relationships consistent with the empirical data.

Semiparametric AFT partial linear models have been proposed to incorporate the non-linear function of a continuous covariate [27, 42] with the following specification:

$$\log T = \boldsymbol{\beta}^T \boldsymbol{X} + f(Z) + \varepsilon \tag{2.5.1}$$

where Z is a 1-dimensional covariate, and the function $f(\cdot)$ is an unknown function that represents a smooth, potentially non-linear relationship between the covariate Z and the log event time.

The concept of partial linear models was firstly introduced in classic linear regression. Heckman and Rice proposed to estimate $f(\cdot)$ using splines, and Speckman [84] using kernel smoothing techniques. Orbe et al. and Zou et al. have further extended the partial linear model in the context of AFT model that estimates $f(\cdot)$ with natural cubic splines and penalized splines [27, 42], based on weighted least squares estimation and rank-based estimation, respectively. The error distribution is left unspecified in both approaches. Nevertheless, the estimation of the baseline hazard, as well as the associated survival probability, was not discussed. Moreover, both of the methods only studied the case where the covariate is a univariate component while assuming the time ratio is constant over time. Methods that can incorporate non-linear structures for all possible continuous variables are desirable for real-life multivariate analysis.

2.5.2 The constant time ratio assumption in the AFT model

The conventional AFT model assumes that the time ratio is constant over the study duration. The relationship between the conditional survival function and the baseline survival function, $S(t|\boldsymbol{X}) = S_0\big(\exp(\boldsymbol{\beta}^T \boldsymbol{X})t\big)$, implies that the ratio of two survival times is constant for any given survival probabilities [3, 8]. Specifically, for a binary treatment covariate, the proportion of patients who have survived in the treatment group at any time point t_1 is the same as the proportion of those who have survived in the control group at any time t_0, where $t_0 = \exp(\beta)t_1$. For instance, if the estimated time ratio is 0.8, the median time for the control group is 0.8 times the median time for the treatment group, and this relative ratio holds true for any percentile

of the survival distribution [72]. This assumption can be violated if the time ratio varies over time, depending on what percentile of the survival time is being evaluated.

The literature on time-dependent time ratios from the AFT model is limited. To the best of my knowledge, it is only briefly discussed by Orbe when comparing PH and AFT models for survival analysis [9]. In addition, only one single paper that proposes a method of modeling the time-dependent time ratio is identified. This method is developed by Elsayed, Liao and Wang [44] based on Etezadi-Amoli and Ciampi's extended hazard regression (EHR) model with the following specification:

$$\lambda(t|\boldsymbol{X}) = \exp\left((\alpha_0 + \alpha_1 t)\boldsymbol{X}\right)\lambda_0\Big(\exp\left((\beta_0 + \beta_1 t)\boldsymbol{X}\right)t\Big) \tag{2.5.2}$$

where the baseline hazard function $\lambda_0(\cdot)$ is a quadratic function:

$$\lambda_0(\cdot) = \gamma_0 + \gamma_1 t + \gamma_2 t^2. \tag{2.5.3}$$

This model has a very generic form that encompasses both Cox PH and AFT models, as well as their extensions for time-dependent effects, as special cases. However, it assumes that the time ratio is changing linearly with time, and the baseline hazard has a quadratic shape, two assumptions that are not always realistic in real-world applications. Moreover, further development is required to take into account the possible non-linear functions of all continuous covariate effects and to accommodate more flexible functions for the time-dependent time ratios for all covariates included in the AFT model.

2.5.3 Diagnostics for AFT models

Several graphic methods can be used to assess whether the AFT model is appropriate concerning the constant time ratio assumption and whether a particular parametric event time distribution fits the observed data well.

Quantile-quantile (QQ) plot

The AFT constant time ratio assumption can be checked by the so-called quantile-quantile (QQ) plots for variables with two levels. It follows that for any quantile q of the survival time distribution, where $0 < q < 1$ and $1 - q = S_0(t_0) = S_1(t_1)$, we have that $\frac{t_0^{(q)}}{t_1^{(q)}} = e^{\beta}$. For a univariate setting with a binary covariate, we can compute the Kaplan-Meier curves to estimate the quantiles for the two groups (e.g., treated vs. control), i.e., $t_0^{(q)}$, $t_1^{(q)}$, for various values of q. If the constant time ratio assumption holds, plotting the estimated percentiles in the control group ($t_0^{(q)}$) against the estimated values ($t_1^{(q)}$) in the treatment group would yield an approximately straight line through the origin [3, 52]. The slope of this straight line should be close to the estimate of e^{β}.

Cox-Snell residual plot

Cox and Snell proposed a type of residual for censored data that can be used to assess the goodness-of-fit of a general survival model [85]. For the i^{th} individual, the residuals are defined as:

$$r_i = -\log\left(\hat{S}(T_i|\boldsymbol{X}_i)\right) = \hat{H}(T_i|\boldsymbol{X}_i). \tag{2.5.4}$$

The general idea behind the Cox-Snell residual is that the survival function $S(T_i|\boldsymbol{X}_i)$ follows a uniform $U[0, 1]$ distribution, and thus $-\log\left(\hat{S}(T_i|\boldsymbol{X}_i)\right)$ has a unit exponential distribution. When the i^{th} individual is censored, the corresponding residual is censored too. If the model being assessed is correct, the Cox-Snell r_i's should represent a censored sample from a unit exponential distribution. We can then plot the Nelson-Aalen estimator of the cumulative hazard function of the residuals $\hat{H}(r_i)$ against r_i. A straight line through the origin with a slope of 1 indicates an adequate fit to the data.

Statistical criteria

To compare the goodness-of-fit of a variety of candidate models for survival data, the Akaike information criterion (AIC) can be used [86]. The AIC for a particular model is

calculated as follows:

$$AIC = 2(df - \hat{\ell}) \tag{2.5.5}$$

where df is the number of estimated parameters in the model, and $\hat{\ell}$ is the maximum likelihood of the model. While AIC rewards goodness-of-fit, the number of degrees of freedom (df) including the number of covariates in the model and the number of additional 'ancillary' parameters in the parametric or semiparametric model, is considered as a penalty for model complexity. Although there is no formal test to compare different AIC values, the model with a lower value is generally preferable.

Similar to AIC, the Bayesian Information Criteria (BIC) for survival data is defined as [87]:

$$BIC = \log(n) \times (df - \hat{\ell}) \tag{2.5.6}$$

where n is the number of uncensored events. BIC places a larger penalty for model complexity compared to the AIC, which discourages overfitting more strictly. It has been suggested that BIC is more appropriate for model selection when the 'true model' is in the candidate set, whereas AIC tends to have better practical performance over BIC to find the best-approximating model if the 'true model' is not in the candidate set [88, 89].

2.6 Parametric AFT models

The parametric AFT model requires the complete specification of the event time (T) distribution. Common parametric choices include the exponential, Weibull, log-logistic, log-normal, and gamma distributions, etc. The event time distribution inherently determines the distribution of the random term W in equation (2.3.1), as summarized in Table 2-1. The estimations and inferences of these parametric AFT models and the corresponding proprieties are discussed in [3, 7, 54], and their survival functions, hazard rates, probability density

functions are summarized in Table 2.2 (page 38) of the book by Klein and Moeschberger [52].

Table 2-1: Common parametric AFT models with corresponding distributions of T and W

Distribution of T	Distribution of W
Exponential	Extreme value
Weibull	Extreme value
Log-logistic	Logistic
Log-normal	Normal
Gamma	Log-gamma

As mentioned in section 2.4, the Weibull distribution, which reduces to the exponential distribution as a special case, is the only continuous distribution that yields both a PH and AFT model. Specifically, when the event distribution is Weibull, the baseline hazard function in (2.3.7) is given by $\lambda_0(t) = \alpha\lambda^{\alpha-1}$, where α denotes the shape parameter. This specification is equivalent to the parameterization that sets $\lambda = \exp(-\frac{\mu}{\sigma})$ and $\alpha = \frac{1}{\sigma}$ in (2.3.2) with ε follows a standard extreme value distribution with the probability density function $f(\varepsilon) = \exp(\varepsilon - e^{\varepsilon})$. For a given individual with covariate vector $\boldsymbol{X}$, the hazard function is given by,

$$\lambda(t|\boldsymbol{X}) = \exp(\boldsymbol{\beta}^T\boldsymbol{X})\lambda_0\big(\exp(\boldsymbol{\beta}^T\boldsymbol{X})t\big) = \exp(\boldsymbol{\beta}^T\boldsymbol{X})\Big(\alpha\lambda\big(\exp(\boldsymbol{\beta}^T\boldsymbol{X})t\big)^{\alpha-1}\Big). \tag{2.6.1}$$

Rescaling $\boldsymbol{\beta} = \frac{\boldsymbol{\beta}^*}{\alpha} = \boldsymbol{\beta}^*\sigma$ leads to the following PH model:

$$\lambda(t|\boldsymbol{X}) = \exp(\boldsymbol{\beta}^{*T}\boldsymbol{X})\lambda_0(t) = \exp(\boldsymbol{\beta}^{*T}\boldsymbol{X})(\alpha\lambda t^{\alpha-1}). \tag{2.6.2}$$

The relationship between (2.6.1) and (2.6.2) facilitates the translation of the covariate effect from the time ratio to the HR in the Weibull AFT model [52]. Specifically, if we model the time ratios, $e^{\boldsymbol{\beta}}$, in the Weibull AFT model, the PH-based hazard ratio can be derived from the same model corresponding to HR=exp$(\frac{\boldsymbol{\beta}}{\sigma})$. In real-life analyses, if a large discrepancy is revealed between the HR converted from the Weibull AFT model and the one estimated directly from the Cox PH model, the underlying event time distribution is likely not Weibull.

On the other hand, if the event times indeed follow a Weibull distribution, we would gain power by modeling the data with the correct parametric assumption.

It is also interesting to note that the log-logistic model is the only parametric model that can be represented as both an AFT and a proportional odds (PO) model [52]. The baseline survival function from a log-logistic distribution is given by:

$$S_0(t) = \frac{1}{1 + \lambda t^{\alpha}}. \tag{2.6.3}$$

The specification is equivalent to parametrizing $\lambda = \exp(-\frac{\mu}{\sigma})$, and $\alpha = \frac{1}{\sigma}$ in (2.3.2) with ε follows a standard logistic distribution with probability density function, $f(\varepsilon) = \frac{e^{\varepsilon}}{(1+e^{\varepsilon})^2}$. Under the AFT model, the conditional survival function given $\boldsymbol{X}$ is,

$$S(t|\boldsymbol{X}) = S_0\big(\exp(\boldsymbol{\beta}^T \boldsymbol{X})t\big) = \frac{1}{1 + \lambda\big(\exp(\boldsymbol{\beta}^T \boldsymbol{X})t\big)^{\alpha}}. \tag{2.6.4}$$

Rescaling $\boldsymbol{\beta} = \frac{\boldsymbol{\beta}^*}{\alpha}$ yields:

$$S(t|\boldsymbol{X}) = \frac{1}{1 + \lambda \exp(\boldsymbol{\beta}^T \boldsymbol{X})t^{\alpha}}. \tag{2.6.5}$$

such that for a binary treatment covariate X, we have the odds of survival beyond time t following a specification of a logistic model:

$$\frac{S(t|X=1)}{1 - S(t|X=1)} = \exp(-\beta^*)\frac{S(t|X=0)}{1 - S(t|X=0)}. \tag{2.6.6}$$

This relationship allows us to translate the covariate effect from the time ratio to the odds ratio, and $e^{-\beta^*}$ is the relative odds of having the event comparing the treated and control groups.

Diagnostic plots for parametric AFT models

The fit of a parametric AFT model can be assessed through properties implied by the parametric distributional form of the event time. Some useful properties are summarized in the books by Cox and Oakes, and Klein and Moeschberger comparing different parametric families [3, 52]. For example, we can plot (a) $\log \hat{H}(t)$, (b) $\log\big(\exp(\hat{H}(t) - 1)\big)$, or (c) $\Phi^{-1}\Big(1 - \exp(\hat{H}(t))\Big)$ respectively against $\log t$ to check if (a) Weibull, (b) log-logistic AFT, or (c) log-normal model fit the data adequately. An approximate straight line indicates

a reasonable fit. Φ^{-1} is the inverse of the standard normal cumulative distribution function, and $\hat{H}(t)$ can be obtained by the Nelson-Aalen estimator described in section 2.1. Similarly, a straight line plotting $\hat{H}(t)$ against t indicates that the observed event time follows an exponential distribution.

It is also noted that a graphical assessment of goodness-of-fit can be made by plotting $\hat{S}^{(-1)}(\hat{S}_{KM}(t))$ against $\log t$ and checking for linearity [61]. $S(\cdot)$ is the analytical function form of the survival distribution being assessed, and $\hat{S}_{KM}(t)$ is the Kaplan-Meier estimator of survival. Therefore, we can also plot (a) $\log(-\log(\hat{S}_{KM}(t))$, (b) $\text{logit}(\hat{S}_{KM}(t))$, or (c) $\hat{\Phi}^{-1}(\hat{S}_{KM}(t))$ respectively to check for the distributional assumptions imposed by (a) Weibull, (b) log-logistic AFT, or (c) log-normal AFT models. These diagnostic plots allow a visual assessment for the parametric assumptions; however, they are restricted in checking for the adequacy of a given model in the univariate setting [52].

2.7 Semiparametric AFT models

Indeed, an appropriate event time distribution might be difficult to identify in complex real-life studies, as prior substantive knowledge about the prognosis of the disease under study over time is usually limited. Two classical semiparametric approaches include the least-squares-regression-based estimator introduced by Buckley and James [12, 13] and the linear-rank-test-based estimator proposed by Tsiatis [14, 16, 17], where the event time distribution is completely unspecified. However, these semiparametric estimating equations have potentially multiple solutions, leading to computational problems, especially with a large number of covariates [6, 18, 19]. In addition, most of these estimators focus on the inference about the covariate effects, providing little theories and tools for the estimations of the hazard or survivor functions that describe the temporal behavior of the event times. Park and Wei [90] derived an estimator of subject-specific survival functions based on the covariate effect estimates from the rank based methods and the Nelson-Aalen estimator of the cumulative hazard function, nevertheless the

implementation of this procedure is not available in standard statistical software. A single summary of the relative covariate effect, e.g., HR or time ratio, does not provide adequate information regarding the impact of the treatment or prognostic factor on survival, as the same relative risk implies different absolute risk difference depending on the baseline hazard rate [55]. Cox and Oakes [3] noted that "it may be physically enlightening to consider the immediate 'risk' attaching to an individual known to be alive at age t; comparisons of groups of individuals are sometimes most incisively made via the hazard." Therefore, the hazard function and the associated time-specific survival probabilities are important quantities to be considered in survival analysis. A few methods, briefly described in the following subsections, have been developed to accommodate these estimates without making an assumption about the event time or error distribution.

2.7.1 Smoothed error AFT model developed by Komárek et al.

Komárek et al. [26] developed a semiparametric approach for fitting AFT models of the form (2.3.2) with respect to the error distribution ε. In their model, the probability density function for ε is approximated by a mixture (or a linear combination) of a large but fixed number of basis Gaussian densities with pre-fixed mean and variance parameters. Specifically, the error density function is expressed as:

$$f(\varepsilon|\boldsymbol{c}) = \sum_{i=1}^{g} c_i \varphi_{\mu_i,\sigma_0^2}(\varepsilon) \tag{2.7.1}$$

where $\varphi_{\mu_i,\sigma_0^2}(\varepsilon)$ is the i^{th} mixture component formed by a basis Gaussian density function with mean μ_i and common variance σ_0^2, and $\boldsymbol{c} = (c_1, \cdots, c_g)$ are the mixture coefficients. To ensure that $f(\varepsilon|\boldsymbol{c})$ is a valid density function, the constraint $\sum_{i=1}^{g} c_i = 1$ must be imposed. The coefficients $\boldsymbol{c}$ are then reparametrized by coefficients $\boldsymbol{a}$, where $c_i(\boldsymbol{a}) = \frac{\exp(a_i)}{\sum_{l=1}^{g} \exp(a_l)}$, $i = 1, \cdots, g$, to avoid such constraint. Values of μ_i and σ_0^2 are empirically fixed *a priori*, but the parameter vector $\boldsymbol{a}$ and the regression coefficient ($\boldsymbol{\beta}$) are estimated jointly using the penalized maximum

likelihood (2.7.2),

$$\ell_P(y|\boldsymbol{\theta}, \lambda) = \ell(\boldsymbol{\theta}) - q\{\boldsymbol{a}; \boldsymbol{\lambda}\} \tag{2.7.2}$$

where $\boldsymbol{\theta}$ is the vector of all estimable parameters, and $\ell_P(y|\boldsymbol{\theta}, \lambda)$ and $\ell(\boldsymbol{\theta})$ are the penalized log-likelihood and the ordinary log-likelihood, respectively. Penalization with a smoothing parameter λ is applied to the squared difference of the transformed coefficients ($\boldsymbol{a}$) of two adjacent basis Gaussian densities. This tuning parameter λ is used to control the smoothness of the fitted error distribution and is selected by minimizing a cross-validation score given by:

$$CV(\lambda) = -\left\{\ell(\hat{\boldsymbol{\theta}}) - trace(\widehat{\boldsymbol{H}}^{-1}\hat{\boldsymbol{I}})\right\} \tag{2.7.3}$$

where $\widehat{\boldsymbol{H}} = -\partial^2\ell_P(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$, $\hat{\boldsymbol{I}} = -\partial^2\ell(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$, and $trace(\widehat{\boldsymbol{H}}^{-1}\hat{\boldsymbol{I}})$ is called the *effective degrees of freedom*. This approach is considered as an approximation of cross-validation with reduced computation cost. More details on the parameterization and the penalty function are described in Komárek et al. [26]. Two types of variance estimates, i.e., pseudo-variance and asymptotic variance, are derived for making inference for the penalized MLE. Previous studies showed that the pseudo-variance estimate yielded better coverage compared to the asymptotic variance estimate. This method has been implemented in the R package 'smoothSurv' [91].

Once the parameters of the error distribution are estimated, the corresponding hazard and survival function can be reconstructed using their known analytical relationships. However, simulation studies are needed to systematically evaluate the performance of this approach with respect to the accuracy and stability of the corresponding hazard and survival estimations.

2.7.2 Extended hazard regression developed by Ciampi et al.

A general extended hazard regression (EHR) was proposed by Etezadi-Amoli and Ciampi [24, 25], assuming that the covariate affects the hazard function according to:

$$\lambda(t|\boldsymbol{X}) = \exp(\boldsymbol{\alpha}^T\boldsymbol{X})\lambda_0\big(\exp(\boldsymbol{\beta}^T\boldsymbol{X})t\big). \tag{2.7.4}$$

The model in (2.7.4) is a generalization of the PH and the AFT model. When $\boldsymbol{\beta} = 0$ it reduces to a PH model, while when $\boldsymbol{\alpha} = \boldsymbol{\beta}$ it corresponds to an AFT model. This approach is appealing because a less restrictive assumption is made concerning the relationship between the covariate and the hazard, yet, in case the generalized form is not needed, the likelihood ratio test (LRT) could be used to selected either PH or AFT model nested within the EHR. In the EHR, instead of eliminating the baseline hazard in the estimation procedure as in Cox's approach, or assuming an explicit parametric function as in parametric AFT models, $\lambda_0(\cdot)$ is approximated by a function defined by a parameter of dimensionality to be determined by data. For instance, a polynomial of degree 2 [24] and quadratic splines [25] have been considered for this approximation. Maximization of the full likelihood, derived from (2.7.4), can be employed to obtain the estimation for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and the parameters used to define $\lambda_0(\cdot)$. However, constraints must be imposed to ensure that $\lambda_0(\cdot)$ remains nonnegative for all t, resulting in a numerically complicated estimation algorithm required by constrained optimization. In addition, both approaches may not offer sufficient flexibility for shapes of hazard function that correspond to complex disease prognosis in health care research. This might be the reason why the authors have suggested fit polynomials of higher degrees and increase the number of knots until no improvement in the fit is obtained [24, 25]. Furthermore, the performance of this method has not been validated through simulation studies, and the statistical package for implementing the method is not available.

2.8 Flexible survival models in alternative modeling frameworks

2.8.1 Flexible extension of the Cox model developed by Wynant and Abrahamowicz

Wynant and Abrahamowicz proposed a flexible extension of the Cox model for estimation of individual-specific survival curves while taking into account potential non-linear (NL) and time-dependent (TD) covariate effects [41]. This method involves a two-stage procedure. In the first stage, a 'product model' (2.8.1) is employed to estimate the NL and TD effects using partial likelihood estimation:

$$\lambda(t|\boldsymbol{X}) = \lambda_0(t) \exp\big(\sum_j \beta_j(t) * g_j(X_j)\big). \tag{2.8.1}$$

$\beta_j(t)$ and $g_j(X_j)$ are included in the model to relax the PH and linearity assumptions imposed by the conventional Cox model for the covariate X_j. Both $\beta_j(t)$ and $g_j(X_j)$ are modeled with low-dimension un-penalized quadratic regression B-splines, whereas the baseline hazard is again treated as a nuisance parameter in this stage as in the conventional Cox PH model:

$$\beta_j(t) = \sum_{q=1}^{Q} b_{q,j} B_q(t), \tag{2.8.2}$$

and

$$g_j(X_j) = \sum_{l=1}^{L} a_{l,j} A_{l,j}(X_j). \tag{2.8.3}$$

In the second stage, the estimated TD and NL covariate effects are substituted into the full likelihood function, in which the logarithm of the baseline hazard is approximated by cubic regression splines (2.8.4), to accommodate more complex shapes than the NL and TD covariate effects.

$$\lambda_0(t) = \sum_{k=1}^{K} \gamma_k B_k(t). \tag{2.8.4}$$

The spline coefficients $\boldsymbol{\gamma}$ are the only parameters to be estimated in this stage, as $\beta_j(t)$ and $g_j(X_j)$ are considered as known fixed functions obtained from the first stage. Maximizing the full likelihood function yields the estimates of $\boldsymbol{\gamma}$. Subsequently, the survival probability at time t for any given covariate pattern can be calculated by the plugging in $\hat{\lambda}_0(t)$ as well as the NL and TD estimates.

$$S(t|\boldsymbol{X}) = \exp\left\{ - \int_0^t \exp\big(\sum_{k=1}^{K} \hat{\gamma}_k B_k(u)\big) \exp\left\{ \sum_j \big[\sum_{q=1}^{Q} b_{q,j} B_q(u)\big] \big[\sum_{l=1}^{L} a_{l,j} A_{l,j}(X_j)\big]\right\} du\right\} \tag{2.8.5}$$

The regression B-splines have been intensively exploited in this method. A regression B-spline over a closed interval $[a, b]$ is represented by a linear combination of the splines of a B-spline basis [92, 93]. These splines are a series of piecewise polynomials of degree p, defined over a set of mutually exclusive intervals. The points that divide the mutually exclusive intervals over $[a, b]$ are referred to as interior knots, and there are $p + 1$ exterior knots on each side of the interval $[a, b]$. With m interior knots and degree p, there are $m + p + 1$ splines. For $j = 1, \cdots, m + p + 1$, each spline basis $B_{j,p}(x)$ is defined recursively as [94]:

$$B_{j,0}(x) = \begin{cases} 1 & \text{if } k_j \le x < k_{j+1} \\ 0 & \text{otherwise} \end{cases}, \tag{2.8.6}$$

and

$$B_{j,p}(x) = \frac{x - k_j}{k_{j+p} - k_j} B_{j,p-1}(x) + \frac{k_{j+p+1} - x}{k_{j+p+1} - k_{j+1}} B_{j+1,p-1}(x) \tag{2.8.7}$$

where the locations of the knots are placed at: $a = k_1 = \cdots = k_{p+1} < k_{p+2} < \cdots < k_{m+p+1} < k_{m+p+2} = \cdots = b$. The regression B-spline function $f(x)$ over the interval $[a, b]$ is constructed as:

$$f(x) = \sum_{j=1}^{m+p+1} \alpha_j B_j(x) \tag{2.8.8}$$

where α_j are the spline coefficients to be estimated from the data.

The smoothness of the resulting B-spline function depends on the degree p, the number, and location of the knots, which should be fixed in advance. For example, $p = 0$ represents step functions that are not continuous at the knots, $p = 1$ represents linear splines which are continuous but not smooth, whereas $p = 2$ and $p = 3$ result in quadratic and cubic splines that are both continuous and smooth [93]. Simulations show that quadratic and cubic regression splines are flexible enough to capture a variety of clinically plausible shapes with realistic complexity [36, 40, 95]. On the other hand, real-life studies have suggested that splines with even one or two knots can be sufficient to represent biologically plausible function forms [96, 97]. The interior knots can be placed at positions where more curvature of the function is expected according to prior information [92]. Nevertheless, when such information is not available, the interior knots may be placed at quantiles of the variable [93, 98, 99]. Alternatively, the number and the location of the knots can be considered as hyperparameters and estimated and compared using goodness-of-fit criteria [99, 100]. However, such an approach can be very complex and computationally intensive [92]. The advantage of regression B-splines is that the estimated curves are only locally affected by outliers, as a B-spline basis is only 'active' over a limited interval (it takes the zero value outside the interval) [94]. In other words, B-splines are less sensitive to local bias compared to other flexible techniques that have global control over the entire range of the data [82].

2.8.2 The Hazard Regression (HARE) model of Kooperberg and colleagues

Kooperberg, Stone and Truong proposed to estimate the conditional log-hazard function directly using polynomial splines [101]:

$$\log\left(\lambda(t|\boldsymbol{X};\boldsymbol{\beta})\right) = \alpha(t|\boldsymbol{X};\boldsymbol{\beta}) = \sum_{j=1}^{p} \beta_j B_j(t|\boldsymbol{X}). \tag{2.8.9}$$

The corresponding log-likelihood can be derived as:

$$\sum_{i=1}^{n} \delta_i \Big(\sum_{j=1}^{p} \beta_j B_j(t|\boldsymbol{X}) \Big) - \int_0^{t_i} \exp\Big(\sum_{j=1}^{p} \beta_j B_j(t|\boldsymbol{X}) \Big) du. \tag{2.8.10}$$

The MLE of $\boldsymbol{\beta}$ is obtained by the Newton-Raphson method, and the estimates of the hazard and survival functions can be computed correspondingly. $B_j(t|\boldsymbol{X})$, $j = 1, \cdots, p$ is the j^{th} spline basis function of the event time, covariates and potentially their tensor products over a p-dimensional ($1 \leq p < \infty$) linear space G on $[0, \infty] \times \chi$ such that $g(\cdot|x)$ is bounded on $[0, \infty]$ for $g \in G$ and $x \in \chi$. For example, for two covariates, the following six basis functions ($p = 6$) may be considered as an allowable space $G : B_1 = 1, B_2 = (1-t)_+, B_3 = x_1, B_4 = (x_1 - 6)_+, B_5 = x_1 x_2$, and $B_6 = x_1(1-t)_+$, where $(t)_+ = \max(t, 0)$. The dimension and the basis functions are selected adaptively via stepwise addition and deletion. In the stepwise addition, at each step, a new basis function is chosen among various candidates including (i) linear terms for variables that are not yet in the model, (ii) tensor product of two basis functions for variables already in the model, (iii) basis functions that based upon a new knot in time, and (iv) basis functions that based upon a new knot in variables already in the model. Subsequently, a stepwise deletion algorithm is applied, and the basis function corresponding to the smallest Wald statistic is deleted at each step. Finally, the BIC is used to select the final model among the sequence of models obtained during the stepwise addition and stepwise deletion process.

In this method, the inclusion of the tensor product of covariate and time yields a non-proportional hazard model, while the tensor product of two covariates allows interactions between them. The conditional hazard function and survival function can be estimated directly without any assumption about the relationship between the covariate and the baseline hazard. However, linear splines used in HARE may yield clinically implausible broken-line estimates, and the robustness of the complex model selection procedure is questioned [102]. In addition, unlike the hazard ratio from the Cox model and time ratio from the AFT model, it is unclear how to interpret the fitted coefficient from the HARE model.

2.8.3 Flexible parametric proportional-hazards (PH) and proportional-odds (PO) models of Royston and Parmar

Royston and Parmar proposed flexible parametric models that are extensions of the Weibull PH and log-logistic PO models [103, 104]. For the PH model,

$$\log[-\log\big(S(t;\boldsymbol{X})\big)] = \log\Lambda(t;\boldsymbol{X}) = \log\Lambda_0(t) + \boldsymbol{\beta}^T\boldsymbol{X}, \tag{2.8.11}$$

and for the PO model,

$$\log[S(t;\boldsymbol{X})^{-1} - 1] = \log O(t;\boldsymbol{X}) = \log O_0(t) + \boldsymbol{\beta}^T\boldsymbol{X}. \tag{2.8.12}$$

The baseline log cumulative hazard $\log\Lambda_0(t)$ and the baseline cumulative odds of failure function $\log O_0(t;\boldsymbol{X})$ are modeled using natural cubic splines $s(t;\boldsymbol{\gamma})$, which are cubic splines constrained to be linear beyond boundary knots. Maximizing the full likelihood gives the estimates for the covariate coefficient $\boldsymbol{\beta}$ and the spline coefficient $\boldsymbol{\gamma}$. The interpretation of $\boldsymbol{\beta}$ under the PH model aligns with the hazard ratio as in the Cox PH model, and under the PO model lines up with the odds ratio as in the log-logistic AFT model. Interactions between the spline basis function as a function of t and the covariate are further introduced to allow hazard ratio and odds ratio be to time-varying. No assumption is made for the baseline cumulative hazard or the baseline cumulative odds, so this spline-based approach can potentially accommodate any 'arbitrary' event time distribution. However, this method is proposed within the PH and PO framework, and thus the accelerated failure time interpretation is lost. Simulation studies are needed to assess the performance of this method in terms of covariate effects and hazard functions.

2.9 Model performance in survival prediction

Besides investigating the importance of the effects of the relevant risk factors, one of the primary goals of building a survival model in a prognostic study is to predict survival or risk

at a given follow-up time and to inform patients on their prognosis based their characteristics. Before the implementation of the prediction model in clinical practice, it is important to validate the model and assess the model performance. When several alternative prediction models are considered in the model building process, we may compare the model performance to establish the final model, supplemented by diagnostics of the model assumption and goodness-of-fit discussed in section 2.5.3 and section 2.6. Indexes of survival model's predictive performance can be summarized into three main aspects: (1) overall predictive performance, (2) discrimination, and (3) calibration.

Overall predictive performance

The overall model performance measures how well the model predicts the observed outcome. The prediction error $(Y - \hat{Y})$ is one of the most common indexes for continuous outcomes, whereas the Brier score is used for binary outcomes calculated by the squared differences between the observed binary outcomes and the predicted probability p: $(Y - p)^2$. For time-to-event outcomes, the Brier score is defined based on the time-dependent predictive errors [105]:

$$BS(t) = \frac{1}{n} \sum_{i=1}^{n} \left(I(t_i > t) - \hat{S}(t|\boldsymbol{X}_i) \right)^2 \tag{2.9.1}$$

where $I(t_i > t)$ is the observed event status, and $\hat{S}(t|\boldsymbol{X}_i)$ is the predicted event-free probability at time t conditional on covariate vector X. For survival probability predicted at fixed time points, the Brier scores can be calculated correspondingly. It is also possible to incorporate the inverse probability of censoring weights in the Brier score to account for informative censoring [106]. The Brier score can range from 0 for a perfect prediction to 0.25 for a random prediction with a 50% outcome incidence.

In addition, the global likelihood ratio statistic, defined as $LR = -2(\ell_0 - \ell)$, can be useful for assessing whether the predictors are associated with the response. ℓ_0 and ℓ denote the log-likelihood of the null model and the fitted model, respectively. Analogous to R^2 in an

ordinary model, this quantity can be considered as the -2 log-likelihood 'explained' by the model. In the survival model, for a sample size n, Nagelkerke's R^2, ranging from 0 to 1, is defined based on the likelihood ratio statistics [107]:

$$R_N^2 = \frac{1 - \exp\left(-\frac{LR}{n}\right)}{1 - \exp\left(1 - \frac{2\ell_0}{n}\right)}. \tag{2.9.2}$$

This index is easy to calculate and has been shown to perform well to measure the predictive ability of a Cox PH model [108, 109].

Discrimination performance

Discrimination describes how well the prediction from the model discriminates those with and those without the outcome. The C-statistic, analogous to the area under the 'receiver operating characteristic' (ROC) curve (AUC), is the most commonly used discrimination performance index for a binary outcome. It is a probability of concordance between predicted and observed outcome, with $c = 0.5$ for random prediction and $c = 1$ for a perfectly discriminating model. Harrell et al. proposed a c-statistic for censored survival data [110, 111]:

$$\frac{\sum_{i \neq j} \delta_i I(t_i < t_j) I(\hat{\boldsymbol{\beta}} \boldsymbol{X_i} > \hat{\boldsymbol{\beta}} \boldsymbol{X_j})}{\sum_{i \neq j} \delta_i I(t_i < t_j)}. \tag{2.9.3}$$

It measures the fraction of all pairs of subjects such that the subject with the higher prognostic score (e.g., the linear predictor $\hat{\boldsymbol{\beta}} \boldsymbol{X}$ from the survival fitted model) is the one who survived longer. This calculation is based only on the 'useable' pairs whose survival time can be ordered. For example, if both subjects in a pair are censored or if one has censored, but the censoring time is less than the even time of the other, they do not contribute to the Harrell's c-statistics. Therefore, this index may depend on study-specific censoring distribution. To address this limitation, Uno et al. derived an estimate of c-statistics truncated at a pre-specified time point τ [112]:

$$\hat{c}_\tau = \frac{\sum_{i=1}^n \sum_{j=1}^n \delta_i (\hat{S}_{km}(t_i))^{-2} I(t_i < t_j, t_i < \tau) I(\hat{\boldsymbol{\beta}} \boldsymbol{X_i} > \hat{\boldsymbol{\beta}} \boldsymbol{X_j})}{\sum_{i=1}^n \sum_{j=1}^n \delta_i (\hat{S}_{km}(t_i))^{-2} I(t_i < t_j, t_i < \tau)} \tag{2.9.4}$$

where $I(\cdot)$ is the indicator function, and $\hat{S}_{km}(\cdot)$ is the Kaplan-Meier estimator for the censoring distribution. This formulation involves weighting the uncensored probability in the estimation, therefore making the estimates censoring independent. A related statistic, Somer's D_{xy} correlation, is defined as $2(c - 0.5)$.

Both Harrell's and Uno's c-statistics provide a global summary of the overall discrimination of risk prediction from a fitted survival model. For the prediction of t-year survival, Uno et al. proposed a time-dependent AUC that can be used to assess the model discrimination at specific time points [113]. A working model: $P(T \leq t) = g(\boldsymbol{\beta X})$ is considered, where $g(\cdot)$ is the cumulative density function of the event time distribution associated with the covariate vector X. For a binary prediction rule indexed by a threshold w: $I(g(\hat{\boldsymbol{\beta}}\boldsymbol{X}) > w)$, the time-specific sensitivity and specificity can be estimated respectively by:

$$\widehat{SE}_t(w) = \frac{\sum_{i=1}^{n} \delta_i I(g(\hat{\boldsymbol{\beta}}\boldsymbol{X_i}) > w, t_i < t)/\hat{S}_{km}(t_i)}{\sum_{i=1}^{n} \delta_i I(t_i < t)/\hat{S}_{km}(t_i)}, \tag{2.9.5}$$

and

$$\widehat{SP}_t(w) = \frac{\sum_{i=1}^{n} \delta_i I(g(\hat{\boldsymbol{\beta}}\boldsymbol{X_i}) \leq w, t_i > t)}{\sum_{i=1}^{n} \delta_i I(t_i > t)}. \tag{2.9.6}$$

The ROC at time t can be constructed based on $\{[1 - \widehat{SP}_t(w), \widehat{SE}_t(w)], 0 \leq w \leq 1\}$, and subsequently, the Uno's AUC(t) is calculated by integrating the area under the ROC(t) curve. Other similar time-dependent AUC statistics include methods based on conditional Kaplan-Meier estimator and Nearest neighbor methods [114]. However, they are derived under the assumption that the working model is correctly specified, whereas Uno's approach does not require this assumption [113].

The G-index is another measure of a model's discrimination performance. It computes the mean absolute difference of the prognostic score over all possible pair of subjects:

$$g = \frac{1}{n_{i \neq j}} \sum_{i \neq j} |\hat{\boldsymbol{\beta}}\boldsymbol{X_i} - \hat{\boldsymbol{\beta}}\boldsymbol{X_j}| \tag{2.9.7}$$

Since the G-index is based on the linear predictor only, it can be applied generally to both parametric and semiparametric survival models and is independent of censoring and other complexities [61].

Calibration performance

Calibration refers to the agreement between observed outcomes and predictions [115]. A simple calibration measure is calibration-in-the-large, which compares the average predicted outcome with the average observed outcome. In the survival context, to study the calibration of a prognostic model for a fixed time point t, we can group subjects with similar predicted t-year survival probabilities and compare the average prediction to the observed probability obtained from Kaplan-Meier estimates within each group [61]. Harrell suggests using at least 50 subjects per group, although such grouping is arbitrary and may be inaccurate [116]. Calibration curves can be constructed with the prediction from the model on the x-axis and the observed outcome on the y-axis, with a 45° line indicating perfect prediction. This plot can also be considered as a graphical illustration of the Hosmer-Lemeshow goodness-of-fit test, which is another common measure of model calibration for a binary outcome [116]. The Greenwood-Nam-D'Agostino (GND) statistic has been proposed to extend the Hosmer-Lemeshow goodness-of-fit test for survival data without making any specific assumptions for the prediction model [117]. The subjects are split into groups based on the prognostic scores or the predicted values, and the χ^2_{GND} statistic is defined to test whether the average of the predicted survival probabilities agrees with the observed Kaplan-Meier estimates across the groups.

2.10 Model validation

Validation is an essential aspect in the process of developing a predictive model to assess potential overfit of the model, because the objective is to provide valid outcome predictions on future subjects who have not been used to develop the model [115]. Internal validation, also

being referred to as reproducibility, usually involves assessing the validity of the model using subjects in the underlying population that the developed data originated from. On the other hand, external validation is a more stringent process to assess whether the prediction can be generalized to populations that are 'plausibly related' [115]. In the third manuscript of this thesis, we focus on the internal validation to determine the likely performance of the predictive model on new subjects, considering all the relevant calibration and discrimination indexes. Cross-validation and bootstrap resampling are the two most commonly used techniques for internal validation [61]. Cross-validation requires splitting the data randomly into several subsets. One subset is omitted from the model building process and is used as the testing sample to assess the model performance. The remaining subsets are the training sample in which the model is developed. This process is then repeated for all consecutive subsets, and the performance index is estimated as the average of all assessments. Bootstrap validation involves repeatedly building the model in a bootstrap sample (training sample), and evaluating the performance of the model on the original sample (testing sample). Optimism is a measure of model overfitting, calculated as the difference in the index between the bootstrap sample and the original sample. We can then obtain a bias-corrected or overfitting-corrected estimate of predictive performance by subtracting the average optimism from the final model's apparent performance index [61, 118]. Both techniques provide a useful tool to estimate the likely performance of the final predicted model on future data. The advantage of bootstrap validation over cross-validation is that the optimism-corrected performance estimates are more stable; therefore it is preferable when the study sample size is small [118]. In practice, 100-200 bootstrap is sufficient to obtain robust estimates [115].

CHAPTER 3

Objectives of the thesis

In my PhD thesis, I will try to address the challenges in the introduction. The primary objective of the thesis is to propose and validate a new flexible extension of the conventional accelerated failure time (AFT) model for time-to-event data that avoids some of its restrictive assumptions. Specifically, to fill the gaps identified in the above literature review, the overall goal is to develop a comprehensive multivariable modeling approach, which addresses three challenges in the AFT framework: (i) how to avoid pre-specification of the event time distribution while providing accurate estimates of covariate effects, baseline hazard function and individual survival curve, (ii) how to simultaneously relax the linearity and constant time ratio assumptions implied in the conventional AFT model, and (iii) how to make model building strategies and perform prediction in survival analyses by comparing different modeling approaches, while considering the potential violation of the respective underlying assumptions.

To achieve the above overall goal, the specific methodological objectives of my thesis include:

(i) To develop a new multivariable semiparametric AFT model that allows estimation of arbitrary shapes of the baseline hazard function,

(ii) To extend the method developed in (i) to account for potential time-dependent effects and non-linear effects of continuous variables,

(iii) To validate the above two methods in simulations regarding their performance in the estimation of covariate effects, baseline hazard function, and survival curves, conditional on possibly complex covariate effects.

The substantive objective of the thesis is to illustrate the practical usefulness and the potential insights we could gain from the proposed novel models. To this end, I have applied the proposed AFT models, as well as alternative survival models, including the Cox PH model and its flexible extension, to three different real-life clinical studies, re-assessing the associations between the respective prognostic factors and mortality in patients with, (i) colon cancer, (ii) septic shock, and (iii) non-small cell lung cancer.

CHAPTER 4

4.1 Preamble to Manuscript 1

Conventional accelerated failure time (AFT) model requires specifying the distribution for the event time. An elegant semiparametric approach proposed by Komárek et al. can estimate covariate effects while leaving the event distribution unspecified and permitting the estimation of the error distribution, and the inter-related hazard and survival functions. Nevertheless, the accuracy of these estimators has not been systematically evaluated. This manuscript proposes a new spline-based semiparametric AFT model, as an alternative to the Komárek et al. smoothed error method, and compares their performances in comprehensive simulation studies under various data generating mechanisms for the event time distribution. It extends the conventional AFT model by approximating the baseline hazard using regression B-splines. The methodological development in Manuscript 1 provides foundations for the method proposed in Manuscript 2, which aims to account for non-linear and time-dependent effects in the AFT framework, in addition to flexible hazard estimation. The spline-based AFT model developed in Manuscript 1 is also applied as one of the alternative modeling strategies compared in Manuscripts 2 and 3 for survival analysis in real-life prognostic studies.

In this manuscript, I adapt a two-step alternating conditional estimation (ACE) algorithm to maximize the full log-likelihood for estimating the parameters in the proposed spline-based AFT model. Moreover, I develop a dedicated statistical program to implement this new estimation algorithm. Simulation studies are then designed to evaluate this approach under

two scenarios, including a simple setting where the true hazard function follows parametric distributions and another setting where a more complex shape of the true hazard is present. The results show that it provides accurate estimates of the covariate effects, hazard function, and survival curves in both scenarios and yields more stable hazard estimates comparing to the Komárek et al. method. Based on the satisfactory performance validated in Manuscript 1, the ACE algorithm and its implementation are further extended for estimation of the more complex model developed in Manuscript 2.

The main results of Manuscript 1 have been presented at the 39th Annual Conference of the International Society for Clinical Biostatistics in Melbourne, Australia, 2018, and the Joint Statistical Meetings in Vancouver, Canada, 2018. This article is currently under revision at *Statistics in Medicine*. The publications cited in Manuscript 1 are listed in the reference section at the end of this thesis. Appendix A provides additional information, originally included in the online supporting information of the article submitted to *Statistics in Medicine*, including more details on the estimation algorithm and additional results of the simulation studies and real-life analyses.

Manuscript I: Spline-based Accelerated Failure Time Model

Menglan Pang[1], Robert Platt[1,2,3], Tibor Schuster[4], Michal Abrahamowicz[1,3]

[1]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

[2]Department of Pediatrics, McGill University, Montreal, Quebec, Canada

[3]The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

[4]Department of Family Medicine, McGill University, Montreal, Quebec, Canada

Abstract

The accelerated failure time (AFT) model has been suggested as an alternative to the Cox proportional hazards model. However, a parametric AFT model requires the specification of an appropriate distribution for the event time, which is often difficult to identify in real-life studies and may limit applications. A semiparametric AFT model was developed by Komárek et al. based on smoothed error distribution that does not require such specification. In this article, we develop a spline-based AFT model that also does not require specification of the parametric family of event time distribution. The baseline hazard function is modeled by regression B-splines, allowing for the estimation of a variety of smooth and flexible shapes. In comprehensive simulations, we validate the performance of our approach and compare with the results from parametric AFT models and the approach of Komárek. Both the proposed spline-based AFT model and the approach of Komárek provided unbiased estimates of covariate effects, baseline hazards, and survival curves for a variety of scenarios in which the event time followed different distributions, including both simple and complex cases. However, our proposed spline-based AFT model yielded more stable estimates of the hazard function. As expected, the baseline hazard and survival probabilities estimated by the mis-specified parametric AFT models deviated from the truth. We illustrated the application of the proposed model in a study of colon cancer.

Key words: accelerated failure time model, spline-based method, model misspecification, survival analysis, simulations

4.2 Introduction

Time-to-event analyses are essential to assess the effects of covariates, exposures or interventions on clinical outcomes. Such studies often aim to estimate the survival probability over time for subjects with different values of prognostic factors. However, to accurately estimate

conditional survival functions, the data analyst has to correctly specify both (i) the baseline hazard and (ii) the way particular covariates affect the hazard.

The Cox proportional hazards (PH) model [1] is the most widely used model for survival analysis, with more than 50,000 citations to date. This model permits assessing covariate effects without any assumptions about the distribution of the event time, thus, avoiding the challenges in point (i). However, the PH assumption restricts the estimated HR's to remain constant during the entire follow-up period. If the PH assumption does not hold, the Cox model may lead to misleading conclusions [29] and biased estimates [33]. Indeed, violations of the PH assumption have been frequently reported [35, 45, 119].

The Accelerated Failure Time (AFT) model provides an alternative to the PH model [6, 9, 70]. In the AFT model, the covariates act directly on the time scale, so that the time to event is accelerated or decelerated depending on the covariate value [7, 52]. Yet, parametric AFT models require complete specification of the event time distribution [54], which may affect the estimates of both survival function and covariate effects [10, 11]. Common parametric choices include the log-normal, log-logistic, exponential and Weibull distributions [7, 52]. In the two latter cases, the AFT and the PH models are both valid [54, 120]. However, in many applications, an appropriate event time distribution may be difficult to identify. Whereas one may choose one among a pre-specified set of parametric distributions, using e.g., the Akaike information criterion (AIC) and/or diagnostic plots [52, 72], such data-dependent choices complicate inference. Furthermore, the true baseline hazard function may not follow conventional parametric distributions, including for example non-monotone, asymmetric shapes reflecting abrupt changes in mortality after cancer diagnosis [97]. To avoid difficulties in specifying the parametric baseline hazard model, different semiparametric AFT models were proposed, including least squares regression [12, 13, 14], rank-based estimators [15, 16, 17], weighted least-absolute-deviations method [20], and semiparametric median regressions [121, 122]. Some of these methods are prone to computational problems, especially with many covariates [6]. Furthermore, the aforementioned methods primarily focus on

estimating covariate effects without explicitly deriving the estimators of the hazard or survivor function, although they could be potentially derived by incorporating the estimated error distribution from the Kaplan-Meier method or the Nelson-Aalen estimator of the cumulative hazard function [90].

We identified a few flexible approaches for modeling hazard functions within the AFT framework. In 1980's Etezadi-Amoli and Ciampi [24, 25] proposed a general flexible model for survival analysis, that included both PH and AFT models as special cases. However, complex constrained optimization is required to ensure that the baseline hazard is nonnegative, resulting in numerical problems in the estimation [25]. Moreover, software implementation was not discussed [24, 25]. More recently, Komárek et al. proposed a semiparametric smoothed error AFT model, which relies on a linear combination of a large number of gaussian densities with penalization [26]. This approach does not require specifying the event time distribution, while allowing for the estimation of the hazard and survival functions conditional on any covariate pattern [26]. However, to the best of our knowledge, the resulting hazard and survival estimators has not been yet systematically evaluated, through simulations.

In this article, we propose an alternative flexible AFT model, that employs low-dimension un-penalized regression splines to estimate the baseline hazards of arbitrary shapes, covariate effects, and the survival curves, conditional on covariates. Simulation studies are conducted to validate the proposed spline-based model and compare its performance with conventional parametric AFT models and the semiparametric approach by Komárek et al. [26], under different event time distribution. The spline-based AFT model is then applied to re-assess mortality in colon cancer.

4.3 Methods

4.3.1 General framework of the accelerated failure time model

In the AFT model, the natural logarithm of the event time, $\log T$, is modeled as a linear function of the covariates $X_1, X_2, \cdots, X_J$ [7]:

$$\log T = -(\beta_1 X_1 + \beta_1 X_1 + \cdots + \beta_J X_J) + W \tag{4.3.1}$$

where W is a random error term. Parametric AFT models are usually defined by the distribution of the event times, typically assumed to follow Weibull, log-normal, and log-logistic distribution [7]. The specification of the event time distribution inherently determines the error distribution and the hazard function [7].

The parameters $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_J)$ are the adjusted log time ratios, that are assumed to remain constant across the follow-up time. Exponentiating equation (4.3.1) gives,

$$T = \frac{1}{\exp\left(\sum_{j=1}^{J} \beta_j X_j\right)} T_0 \tag{4.3.2}$$

with $T_0 = e^W$ representing the distribution of the survival times of a hypothetical reference group with 0 values for all measured covariates.

In the AFT model, the hazard function conditional on covariates equals [7]:

$$\lambda(t|\boldsymbol{X}) = \exp\left(\sum_{j=1}^{J} \beta_j X_j\right) \lambda_0\Big(\exp\left(\sum_{j=1}^{J} \beta_j X_j\right) t\Big) \tag{4.3.3}$$

where $\lambda_0(t)$ is the baseline hazard function, corresponding to $X_1 = X_2 = \cdots = X_J = 0$. Accordingly, the survival function at time t, conditional on covariates, is given by:

$$S(t|\boldsymbol{X}) = S_0\Big(\exp\left(\sum_{j=1}^{J} \beta_j X_j\right) t\Big) \tag{4.3.4}$$

Equation (4.3.4) implies that for one unit increase in the value of X_j the time to event is accelerated or decelerated by a multiplicative factor e^{β} [7]. Therefore, the covariate effects estimated from AFT models can be interpreted as the relative prolongation or shortening of the survival time.

4.3.2 Smoothed error AFT model of Komárek et al.

Komárek et al. [26] developed an elegant semiparametric approach for fitting AFT model (4.3.1) without specifying the distributions of event times or errors. They express the density function for the error distribution $f(w|c)$ by a mixture of a large number g of the basis Gaussian densities [26]:

$$f(w|\boldsymbol{c}) = \sum_{i=1}^{g} c_i \varphi_{\mu_i,\sigma_0^2}(w) \tag{4.3.5}$$

where $\varphi_{\mu_i,\sigma_0^2}(w)$ is the i^{th} Gaussian density function with mean μ_i and common variance σ_0^2, and $\boldsymbol{c} = c(c_1, \cdots, c_g)$ are the mixture coefficients. To avoid the need for constrained optimization, and ensure that w meets the conditions for density, the mixture coefficients are reparametrized: $c_i(\boldsymbol{a}) = \frac{\exp(a_i)}{\sum_{l=1}^{g} \exp(a_l)}$, $i = 1, \cdots, g$ so that $c_i(\boldsymbol{a}) = 1$ and $c_i(\boldsymbol{a}) > 0$ [26]. Values of μ_i and σ_0^2 are fixed *a priori*, but the vector $\boldsymbol{a}$ is estimated jointly with the regression coefficients vector ($\boldsymbol{\beta}$) in (4.3.1), using penalized maximum likelihood:

$$\ell_P(y|\boldsymbol{\theta}, \lambda) = \ell(\boldsymbol{\theta}) - q\{\boldsymbol{a}; \boldsymbol{\lambda}\} \tag{4.3.6}$$

where $\boldsymbol{\theta}$ is the vector of all estimable parameters, including $\boldsymbol{a}$,$\boldsymbol{\beta}$ and a scale parameter σ, and $\ell_P(y|\boldsymbol{\theta}, \lambda)$ and $\ell(\boldsymbol{\theta})$ denote the penalized and ordinary log-likelihood function, respectively. The penalty term $q\{\boldsymbol{a}; \lambda\}$ is applied to the squared difference of the transformed coefficients ($\boldsymbol{a}$) of adjacent Gaussian densities $\varphi_{\mu_i,\sigma_0^2}(w)$. The tuning parameter λ, selected by cross-validation, controls the smoothness of the fitted error distribution [26]. Variance of the penalized maximum likelihood estimator is approximated by the pseudo-variance [26]. The Komárek et al.'s smoothed error AFT model is implemented in the R package 'smoothSurv' [91].

4.3.3 Proposed spline-based AFT model

Modeling baseline log hazard by regression splines

We propose an alternative flexible AFT model which does not require penalization. Because the baseline hazard is usually unknown and may be complex, we employ un-penalized low-dimension polynomial regression B-splines that avoid *a priori* parametric assumptions, and approximate well a wide range of functional shapes [92, 93], including density or (log) hazard functions for event times [123, 124]. Therefore, in the proposed flexible AFT model, the log hazard function is modeled as a linear expansion of a basis of $K = p + m + 1$ polynomial regression splines of degree p with m interior knots:

$$\lambda_0\big(\exp\big(\sum_{j=1}^{J}\beta_j X_j\big)t\big) = \exp\big(\sum_{k=1}^{K}\gamma_k S_k(w)\big) \tag{4.3.7}$$

where $w = \exp\big(\sum_{j=1}^{J}\beta_j X_j\big)t$, $S_k(\cdot)$ is the k^{th} B-spline in the basis, and $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_K)$ are the estimable spline coefficients. Modeling the log hazard ensures that the baseline hazard is always positive, and the log-likelihood function is concave, which ensures convergence to a global maximum (second derivatives are shown in Appendix A.1).

Using standard notation $\{t_i, \delta_i, X_{i1}, \cdots, X_{iJ}\}_{i=1}^{n}$ where $t_i = \min(T_i, C_i)$ is the observed time and $\delta_i = I(T_i \leq C_i)$ is the event indicator, the full likelihood for right censored data is given by:

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} = \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S(t_i) \tag{4.3.8}$$

and the full log-likelihood can be expressed in terms of hazard function as follows:

$$\log L = \sum_{i=1}^{n}\Big[\delta_i \log\big(\lambda(t_i|X_{i1}, \cdots, X_{iJ})\big) - \int_0^{t_i}\lambda(u|X_{i1}, \cdots, X_{iJ})du\Big]. \tag{4.3.9}$$

Substituting (4.3.3) and (4.3.7) in (4.3.9), we obtain:

$$\log L = \sum_{i=1}^{n}\Big[\delta_i\Big(\sum_{j=1}^{J}\beta_j X_{ij} + \sum_{k=1}^{K}\gamma_k S_k(w_i)\Big) - \int_0^{t_i}\exp\big(\sum_{j=1}^{J}\beta_j X_{ij}\big)\exp\big(\sum_{k=1}^{K}\gamma_k S_k(w_i)\big)du\Big]. \tag{4.3.10}$$

Estimates of β_j and γ_k are obtained by maximizing the full log-likelihood in (4.3.10), and permit estimating the hazard and survival functions, conditional on covariate vectors. Specifically, conditional survival functions can be computed as:

$$\hat{S}(t|X_1, \cdots, X_J) = \exp\left\{ - \int_0^t \exp\left(\sum_{j=1}^{J} \hat{\beta}_j X_j\right) \exp\left(\sum_{k=1}^{K} \hat{\gamma}_k S_k\left(\exp\left(\sum_{j=1}^{J} \hat{\beta}_j X_j\right) u\right)\right) du \right\} \tag{4.3.11}$$

Alternating conditional estimation

Estimating the parameters of the spline-based AFT model is challenging because the covariates affect the hazard in two different ways: (i) by changing the hazard multiplicatively and (ii) by redefining the baseline hazard function in the time scale, as shown in (4.3.3). (This complication is avoided in the Cox PH model, where the covariate effects are independent of the baseline hazard [1]). Since the same parameter vector $\boldsymbol{\beta}$ (i) needs to be estimated to assess covariate effects on the survival time but also (ii) must be considered as known when estimating spline coefficients $\boldsymbol{\gamma}$ that define the log baseline hazard, we cannot estimate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in a single step. Specifically, it is impossible to derive the score function for the joint log-likelihood, due to the difficulty in obtaining the first derivative with respect to $\boldsymbol{\beta}$ in [7], when it is considered as an unknown parameter in the spline-basis function $S_k(\cdot)$ that needs to be calculated using recursive formulae. To address this complexity, we adapt an iterative alternating conditional estimation (ACE) algorithm [33]. Briefly, the parameter space is divided into two subsets: (i) the $\boldsymbol{\beta}$ vector, and (ii) the $\boldsymbol{\gamma}$ vector. Each ACE iteration involves two steps, and in each step only one of the two subsets of parameters is estimated (i.e. updated) conditional on the most recently estimated values of the other subset, considered at the current step as 'known'. The iterations stop when the difference between the log-likelihoods from two consecutive iterations is less than 10^{-5}. Appendix A.1 provides details of our ACE algorithm. A dedicated R program is also provided for estimating the parameters.

Bootstrap confidence intervals

Standard large-sample inference, based on the covariance matrix of all estimated parameters, does not accurately quantify the sampling variance of the ACE-based estimates [41]. Therefore, we rely on bootstrap [118] to estimate both the 95% confidence intervals (CI) for the covariate effects and the 95% pointwise confidence bands around the survival curves, conditional on covariates. For each of the M bootstrap resamples, we use ACE to estimate the spline-based AFT model (4.3.7). The 2.5^{th} and 97.5^{th} percentiles of the resulting distribution of the M estimates of $\boldsymbol{\beta}$ defines the corresponding 95% CI for the covariate effects. For a given specific covariate pattern, the estimates for both $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are then plugged into equation (4.3.11) to estimate the conditional survival curves $S(t|X_1, \cdots, X_J)$, and 95% pointwise confidence bands are obtained by connecting the 2.5^{th} and 97.5^{th} percentiles of the M corresponding estimates, for each t.

4.4 Simulation Studies

Two simulation studies helped assess the performance of the proposed method under different assumptions about the true hazard. Baseline hazard in simulation B follows a log-normal distribution, but in simulation A represents a mixture of two Weibull distributions, implying a more complex shape than any conventional parametric hazard functions. (Details are described in Appendix A.2).

In both simulations, we generated random samples with $N = 500$. We generated two independent binary covariates (X_1, X_2) from a Bernoulli distribution with probability 0.5, and two continuous, strongly correlated covariates: $X_3 \sim \mathcal{N}(0, 1)$, and $X_4 = X_3^2$. Individual event times were then generated from the following AFT model:

$$\log T = -(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4) + W \tag{4.4.1}$$

where $\beta_1 = 1$, $\beta_2 = -1$, $\beta_3 = \beta_4 = 1$. We assumed a combination of random, uniformly distributed censoring due to losses to follow-up (C_1), and administrative right censoring at $C_2 = 3$ years. C_1 was selected to achieve approximately 25% overall censoring rate in the simulations. The observed time was determined as $\min(T, C_1, C_2)$.

For each simulated scenario, 100 simulated datasets were independently generated and analyzed with each of the following seven alternative models: (i) four conventional parametric AFT models, with different pre-specified baseline distribution: Weibull, exponential, log-normal or log-logistic, (ii) the conventional Cox PH model, (iii) the smoothed error AFT model developed by Komárek et al. [26], and (iv) the proposed spline-based AFT model (4.3.7). In model (4.3.7), un-penalized cubic ($p = 3$) regression B-splines with two ($m = 2$) interior knots, placed at the terciles of the observed follow-up time, were used to estimate the log baseline hazard. In the smoothed error AFT model (iii), we used the default option in the 'smoothSurv' package to select the tuning parameter λ from a grid of values, from $\exp(2)$ to $\exp(-9)$, through cross-validation [26]. In additional analyses, we also forced *a priori* different values of $\lambda = \{\exp(2), \exp(-4), \exp(-8)\}$, in order to assess how the estimates are affected by increasing penalty.

Appendix A.2 describes criteria used to evaluate and compare the performance of different models.

4.5 Simulation Results

4.5.1 Baseline hazard estimates

Figures 4-1 and 4-2 compare the baseline hazard estimates (gray curves) obtained with different estimation models against the true hazard (black dashed curve) for simulations A and B, respectively. Both the proposed spline-based (Figure 4-1(a) and 4-2(a)) and the smoothed error-based (Figure 4-1(b) and 4-2(b)) estimators are free of bias, as the corresponding mean

estimates (white curves) coincide with the true hazard across the follow-up time, with only slight over-estimation in the upper tail where events are sparse. However, individual smoothed error estimates (gray curves in Figures 4-1(b) and 4-2(b)) exhibit large variability, with excessive fluctuation even between $t = 0$ and $t = 2$, when many events occurred. In contrast, spline-based estimates are more stable and accurately reflect the unimodal true shape (Figures 4-1(a) and 4-2(a)).

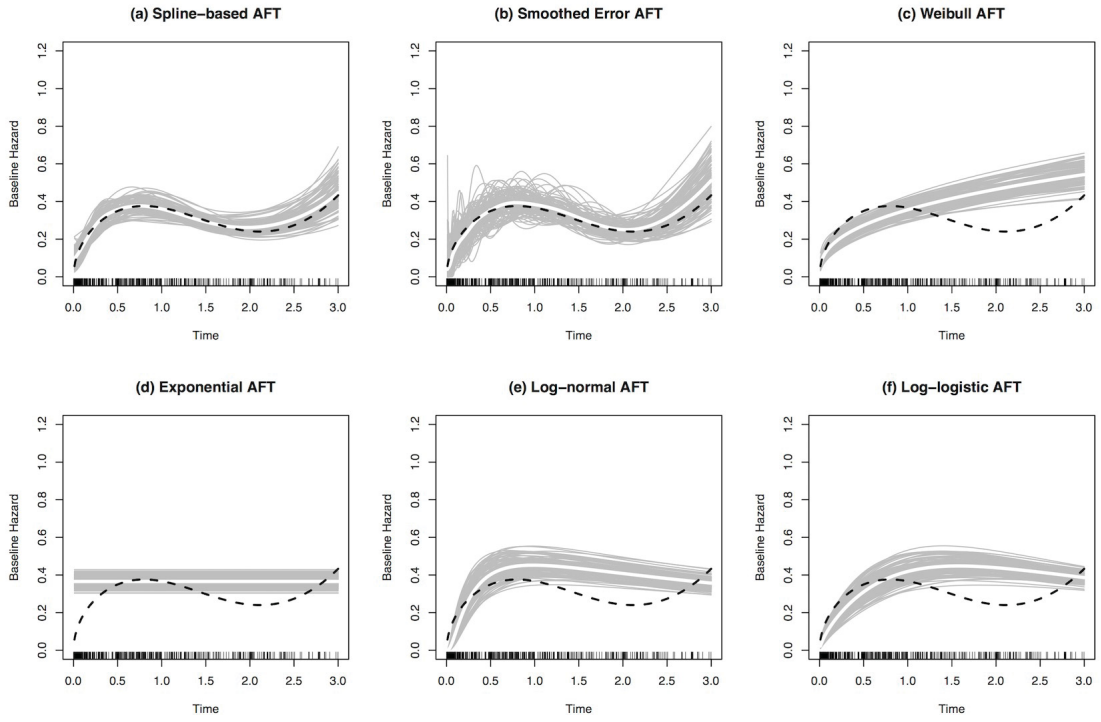


Figure 4-1: Estimated baseline hazard functions using 100 samples in simulation A when the mixture hazard is the true data generating model. The gray curves are the estimated individual baseline hazard functions from 100 samples, and the pointwise mean is shown by the white curve. The black dashed curve represents the true baseline hazard function. The empirical distribution of the observed uncensored event times (75% quantile: 0.88, 90% quantile: 1.58), from one random sample, is shown by rug plot at the bottom of the figures.

As expected, given the complex shape of the true hazard in simulation A, all parametric AFT models yield seriously biased estimates (Figures 4-1(c-f)). In simulation B, the correctly specified log-normal AFT model (Figure 4-2(e)), unsurprisingly, produces unbiased estimates with very good stability. In contrast, all other parametric AFT models, especially the Weibull

and exponential, yield large bias, as they impose constraints inconsistent with the 'true' non-monotone baseline hazard (Figures 4-2 (c-d)).

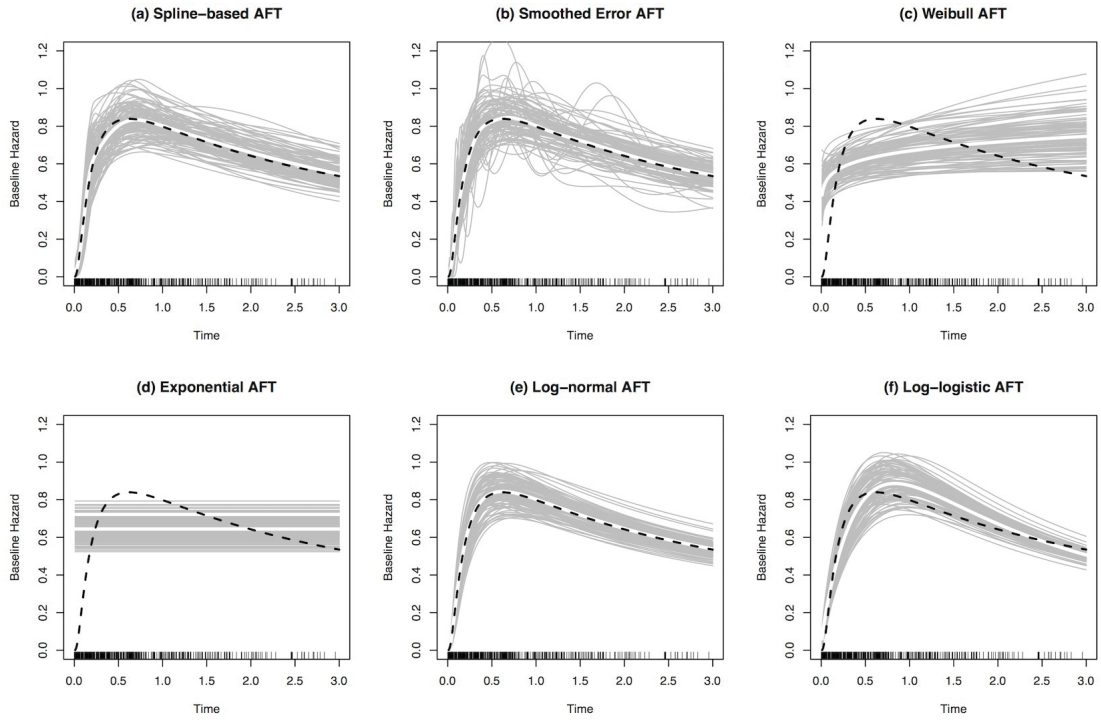


Figure 4-2: Estimated baseline hazard functions using 100 samples in simulation B when the log-normal distribution is the true data generating model. The gray curves are the estimated individual baseline hazard functions from 100 samples, and the pointwise mean is shown by the white curve. The black dashed curve represents the true baseline hazard function. The empirical distribution of the observed uncensored event times, from one random sample (75% quantile: 1.31, 90% quantile: 2.08), is shown by rug plot at the bottom of the figures.

Figure 4-3 summarizes the results of additional analyses that help assess how the smoothed error estimates are affected by the user-specified tuning parameter λ in (4.3.6). In simulation A, with complex true hazard, a weak penalization ($\lambda = \exp(-8)$), similar to the default option, offers unbiased estimates but induces large variance, whereas a strong penalization ($\lambda = \exp(2)$) yields more stable but seriously biased estimates (Figure 4-3(a)). For simulation B, with a simple true log-normal baseline hazard function, there is no bias, regardless of λ, but the estimates become increasingly unstable with lower λ (Figure 4-3(e-f)). Overall, none of the additional λ values we considered performs uniformly better than the default option.

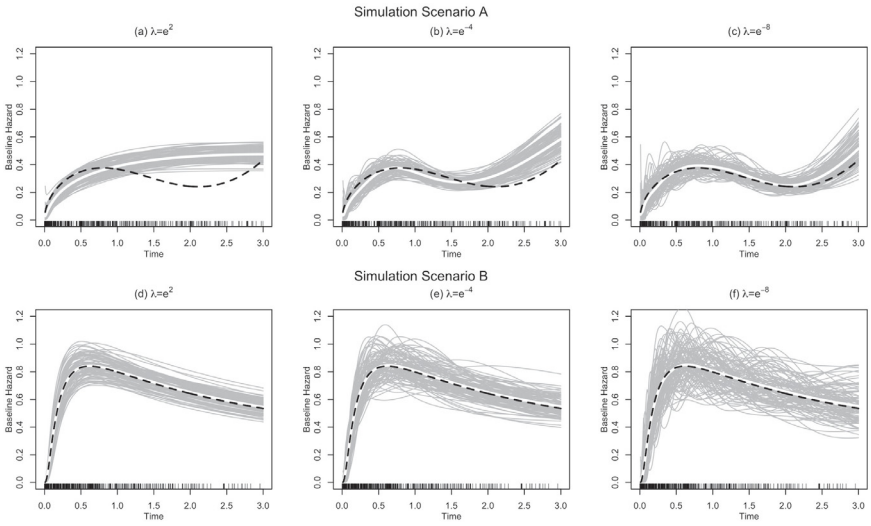


Figure 4-3: Estimated baseline hazard functions using 100 samples in simulation A and B considering different values for the tuning parameter λ in the smoothed error AFT model by Komárek et al. The gray curves are the estimated individual baseline hazard functions from 100 samples, and the pointwise mean is shown by the white curve. The black dashed curve represents the true baseline hazard function.

4.5.2 Estimated covariate effects and AIC comparisons

Table 4-1 compares the accuracy and variability of the estimated covariate effects from the two flexible AFT models. It compares also the AIC from all alternative estimation models, except for the smoothed error model where the effective degrees of freedom are not comparable due to penalization [26]. In simulation A, both smoothed error and spline-based AFT models estimates of the covariate effects are free of noticeable biases and yield consistently smaller root mean squared error (rMSE) than all parametric AFT models (shown in Tables A-1 in the Appendix). As expected, the proposed spline-based AFT model yields much smaller AIC than any of the (mis-specified) parametric models (Table 4-1). In simulation B, the correctly specified log-normal model, as expected, provides unbiased estimates with the smallest rMSE, and the best AIC (Table 4-1). Still, both the proposed spline-based and Komárek et al.'s smoothed error AFT models yield similarly small bias and rMSE and, for the spline-based AFT model, AIC is close to the log-normal model. In contrast, mis-specified Weibull and

exponential models yield somewhat biased covariate effects (shown in Tables A-4 in the Appendix) and the worst AIC's.

Table 4-1: Results of the estimated covariate effects from alternative methods in simulation A and B

		Covariate	Relative Bias† (%)	SD	rMSE	AIC
Simulation A	Spline-based AFT	X_1	-1.1	0.05	0.05	**290.25**
		X_2	-0.1	0.04	0.04	
		X_3	0.2	0.02	0.02	
		X_4	0.1	0.02	0.02	
	Smoothed error AFT	X_1	-0.4	0.04	0.04	299.08*
		X_2	0.3	0.04	0.04	
		X_3	0.5	0.02	0.02	
		X_4	0.2	0.01	0.01	
				AIC		
		Weibull	Exponential	Log-logistic	Log-normal	
		375.14	424.62	449.52	461.18	
Simulation B	Spline-based AFT	X_1	-0.9	0.10	0.10	142.17
		X_2	-1.8	0.12	0.12	
		X_3	0.2	0.05	0.05	
		X_4	0.5	0.04	0.04	
	Smoothed error AFT	X_1	-0.8	0.11	0.11	137.68*
		X_2	-1.0	0.11	0.11	
		X_3	0.3	0.05	0.05	
		X_4	0.3	0.03	0.03	
	Log-normal	X_1	-0.9	0.09	0.09	**137.89**
		X_2	-0.6	0.10	0.10	
		X_3	0.1	0.04	0.04	
		X_4	0.2	0.03	0.03	
				AIC		
		Weibull	Exponential	Log-logistic		
		193.38	198.47	143.95		

†relative bias is defined as $\frac{\hat{\beta}_j - \beta_j}{\beta_j} \times 100\%$

*not comparable to other methods.

4.5.3 Survival curve estimates

Figures 4-4 and 4-5 compare survival curves, estimated with alternative models for an arbitrary covariate pattern ($X_1 = 1, X_2 = X_3 = X4 = 0$), in simulations A and B, respectively. The spline-based and the smoothed error AFT models (Figures 4-4(a-b), Figures 4-5(a-b)) survival estimates are unbiased across the follow-up, reflecting accurate estimation of both the covariate effects (Table 4-1) and the baseline hazard (Figures 4-1 and 4-2). In contrast, most survival curves estimates based on parametric AFT models and PH model (Figure 4-4(c-g), Figure 4-5(c-d, g)) are seriously biased, except for the log-normal AFT model (Figure 4-5(e)) in simulation B where it corresponds to the data-generating model. PH model-based estimates are biased because the underlying assumption is not satisfied. The log-logistic AFT model performs reasonably well in Simulation B (Figure 4-5(f)) because it accommodates the non-monotonic hazard (Figure 4-2(f)). However, the estimated survival probabilities are biased at later times (Table 4-2).

Table 4-2: Relative bias and standard error of the estimated survival probabilities from all the alternative model for different time points in simulation A and B (The results for later time points, e.g., t=2, 2.5, are not provided for simulation A, given that the corresponding true survival probabilities are nearly zero)

		Spline-based AFT		Smoothed error AFT		Weibull AFT		Exponential AFT		Log-normal AFT		Log-logistic AFT		Cox	
	True Survival	Relative Bias† (%)	SD	Relative Bias† (%)	SD	Relative Bias† (%)	SD	Relative Bias† (%)	SD	Relative Bias† (%)	SD	Relative Bias† (%)	SD	Relative Bias† (%)	SD
Simulation A															
t=0.5	0.65	0.70	0.02	-0.23	0.02	4.28	0.03	**-9.84**	0.02	**-6.48**	0.03	1.23	0.03	3.59	0.03
t =1	0.45	-0.77	0.02	-2.58	0.02	**-19.84**	0.03	**-24.03**	0.02	**-22.47**	0.02	**-17.80**	0.03	**-17.19**	0.03
t =1.5	0.17	**5.38**	0.03	-4.78	0.02	0.03	0.02	**19.29**	0.02	**30.97**	0.02	**35.55**	0.02	4.50	0.03
Simulation B															
t=0.5	0.38	1.96	0.04	1.03	0.04	**18.31**	0.04	**13.13**	0.04	0.70	0.04	-1.46	0.04	**14.29**	0.05
t =1	0.16	1.23	0.03	1.13	0.03	**15.08**	0.04	**17.20**	0.04	1.35	0.03	-3.90	0.02	**6.82**	0.04
t =1.5	0.08	2.08	0.02	1.19	0.02	**-9.80**	0.02	1.64	0.02	1.99	0.02	2.38	0.01	**-11.21**	0.03
t=2	0.05	4.78	0.01	1.44	0.01	**-37.62**	0.01	**-20.81**	0.01	2.65	0.01	**13.63**	0.01	**-31.62**	0.01
t =2.5	0.03	2.20	0.01	1.92	0.01	**-60.08**	0.01	**-42.48**	0.01	3.31	0.01	**28.31**	0.01	**-48.47**	0.01
t=0.5	0.38	1.96	0.04	1.03	0.04	**18.31**	0.04	**13.13**	0.04	0.70	0.04	-1.46	0.04	**14.29**	0.05

†relative bias is defined as $\frac{\hat{\beta}_j - \beta_j}{\beta_j} \times 100\%$

Table 4-2 shows the relative bias and empirical standard deviations of the 100 estimates of the probability of survival, for three equidistant time points. Consistent with Figures 4-4 and 4-5, both the spline-based AFT and the smoothed error models perform consistently well.

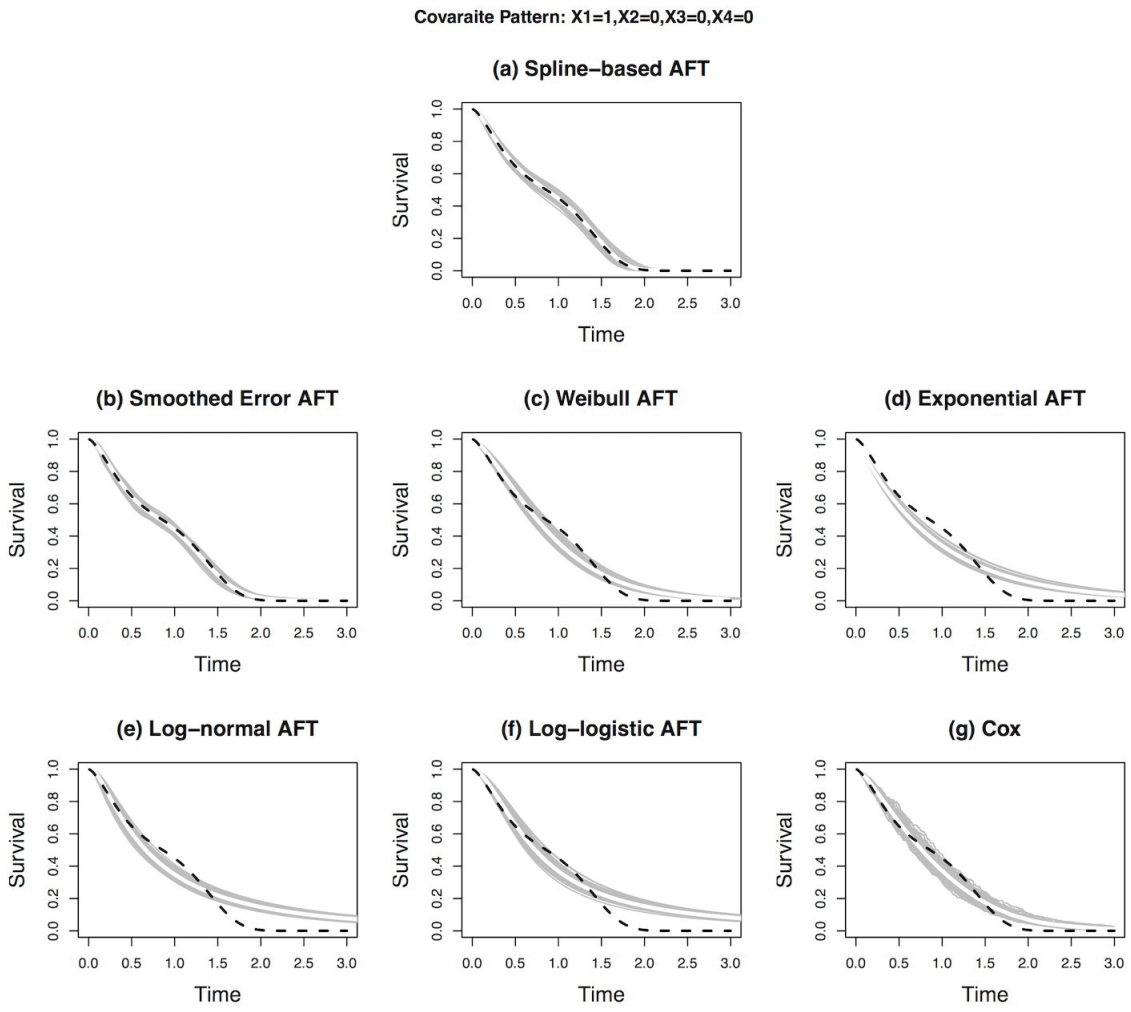


Figure 4-4: Comparison of the survival curve estimates, associated with a specific covariate vector ($X_1 = 1, X_2 = X_3 = X4 = 0$), obtained with alternative estimation models (7 panels) in simulation A when the mixture hazard is the true data generating model. The white curve is the pointwise mean of the estimated individual survival curves from 100 simulated samples (gray curves). The true survival function is represented by the black dashed curve.

In contrast, the PH model and all parametric AFT models, except for the log-normal model in simulation B, yield biased survival estimates for at least some time points. Finally, the variance of the corresponding $S(t)$ estimates is similar for all models (Table 4-2). Similar results were obtained for other covariate vectors (Appendix A.3 and A.4), and similar findings were observed when event times were generated from exponential, Weibull, log-logistic, and gamma distribution (results not shown).

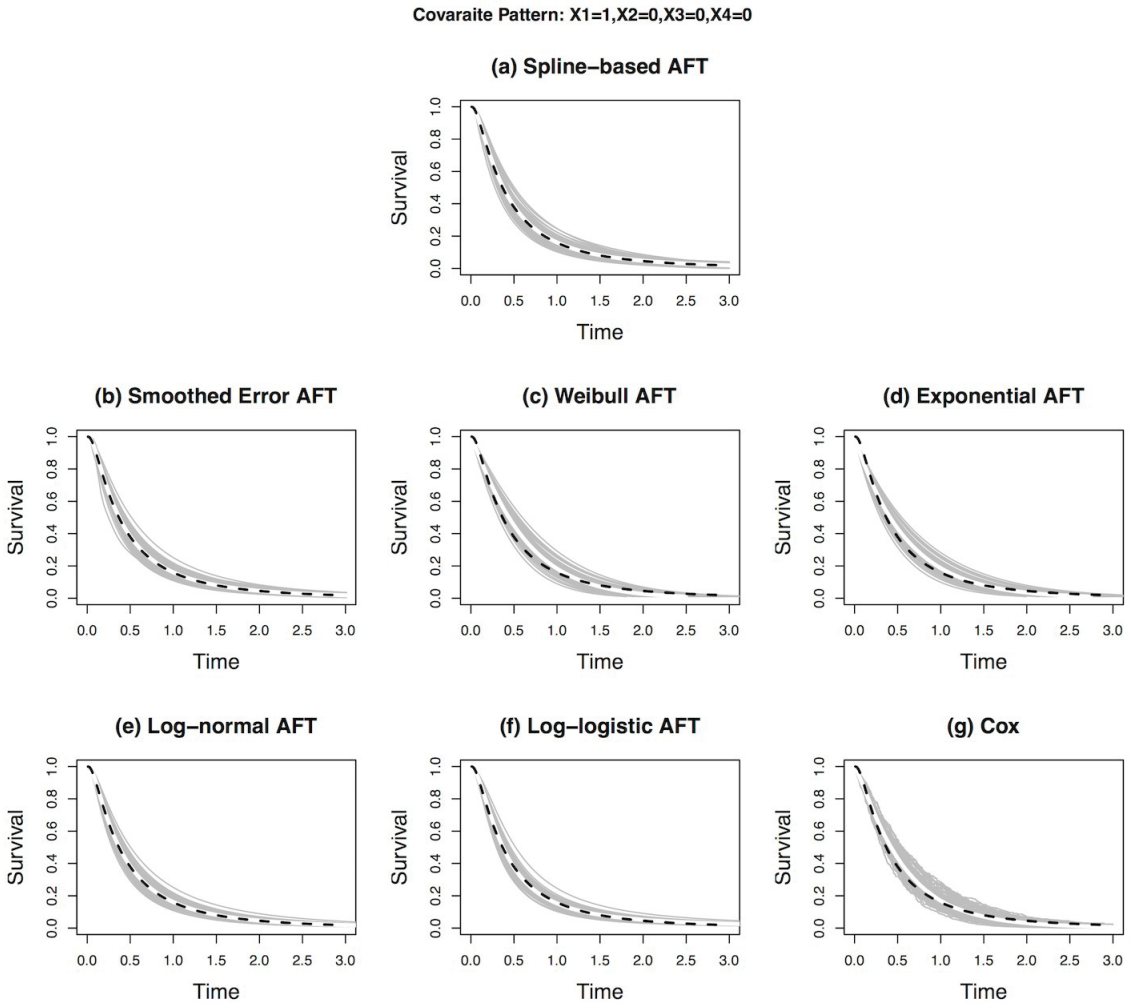


Figure 4-5: Comparison of the survival curve estimates, associated with a specific covariate vector ($X_1 = 1, X_2 = X_3 = X4 = 0$), obtained with alternative estimation models (7 panels) in simulation B when the log-normal distribution is the true data generating model. The white curve is the pointwise mean of the estimated individual survival curves from 100 simulated samples (gray curves). The true survival function is represented by the black dashed curve.

4.6 Application: Survival in colon cancer

Methods

To illustrate a real-life application of the proposed model, we re-analyzed data from a trial of adjuvant chemotherapy in colon cancer. Data are publicly available in the 'survival' R package [71], and described by Moertel et al. [125, 126]. Patients recently diagnosed with stage III

colon cancer, in 1984-87, were enrolled soon after their surgeries, and randomized to: (i) observation only, (ii) levamisole alone, or (iii) levamisole plus fluorouracil (levamisole+5FU). The outcome was time to death of any cause [125, 126]. Baseline covariates included: age, gender, treatment, time since surgery, and several cancer pathological variables: obstruction, perforation, adhesion to nearby organs, histologic differentiation, depth of invasion, and the number of lymph nodes involved. Table A-7 in the Appendix summarizes the covariate distributions. During the median follow-up of 5.4 years, 430 (48%) of the 888 participants died.

All models discussed in Section 4.4 were used to assess the associations of the baseline covariates with all-cause mortality. Un-penalized cubic B-splines, with two interior knots at the terciles of the observed follow-up distribution (3.1 and 6.1 years), were used to implement the proposed spline-based AFT model. The penalization parameter in the smoothed error AFT model was determined by cross-validation [26]. Age was transformed into z-scores, so that the baseline hazard corresponded to the mean age.

Results

The baseline hazards, corresponding to mean age and 0 values of all binary covariates, estimated by different AFT models, are compared in Figure 4-6. The estimates from the parametric AFT models are monotonic, reflecting the underlying distributional assumptions, whereas both more flexible AFT models suggest non-monotone hazards. However, the estimate from the smoothed error AFT model using the default option for the penalty (dashed black curve) seems excessively wiggly, consistent with the simulation results (Figures 4-1 and 4-2). (With increasing penalty, the smoothed error estimates are more similar to the spline-based estimates (Figure A-6 in the Appendix), but the model fit is worse than the penalty chosen by the default option using cross-validation (AIC are shown in Table A-8). In contrast, the proposed spline-based AFT model (solid black curve) yields a much smoother estimate with hazard increasing before 3 years, reaching a plateau from 3 to 4 years and slightly declining thereafter.

The spline-based AFT model fits the data better than all parametric AFT models (Table 4-3), even if the log-normal AFT model has only slightly higher AIC. In the Cox model, the

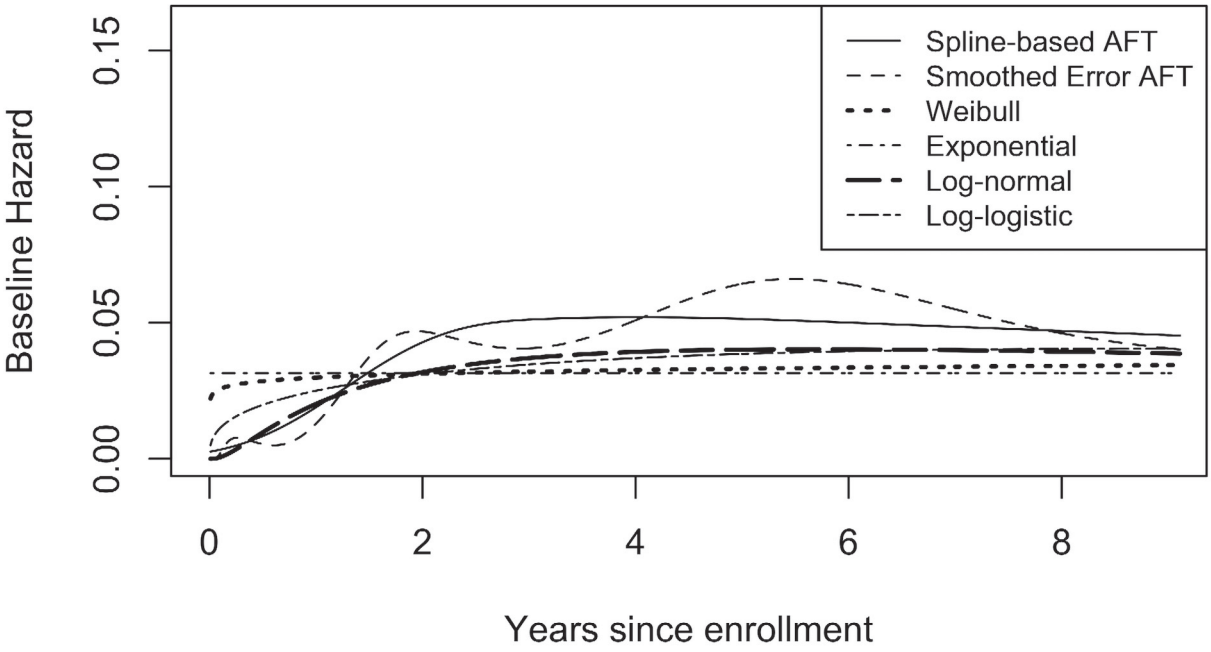


Figure 4-6: The estimated baseline hazards of all-cause mortality in the stage III colon cancer, estimated from different modeling approaches.

Table 4-3: Model fit of alternative analysis methods in the colon cancer study

	Spline-based AFT	Smoothed error AFT*	Weibull	Exponential	Log-normal	Log-logistic
Log-likelihood	-1292.19	-1280.18	-1325.12	-1326.18	-1296.28	-1304.09
df	20	24.69*	16	15	16	16
AIC	2624.38	2609.74*	2682.25	2682.36	2624.57	2640.19

*not comparable to other methods.

global test rejects the PH assumption ($p < 0.001$), which seems violated for obstruction, poor tumor differentiation, and the indicator of positive lymph nodes>4, according to Schoenfeld residuals plot and test (data not shown). On the other hand, the QQ plot (Figure A-7 in the Appendix) does not reveal strong violation of the constant time ratio assumption underlying the AFT models. Furthermore, the Cox-Snell residuals [85] in Figure A-8 suggest that the proposed spline-based and the Komárek's smoothed error AFT models fit the data well, as the

cumulative hazards of the residuals follow the diagonal line, in contrast to important deviations seen in the upper tail for all parametric AFT models and the Cox PH model.

Figure 4-7 shows that alternative AFT models produced generally similar covariate effects (time ratios). (Cox PH model-based hazard ratios are shown in Figure A-9). Based on (i) Cox-Snell residuals in Figure A-8 and AIC comparisons (Table 4-3), below we focus on the results of our proposed spline-based AFT model (the first model in the left panel of Figure 4-7). Levamisole alone and observation only groups have similar survival, but treatment with both levamisole and fluorouracil extends median survival time by about 51.5% (The estimated time ratio (control vs. Levamisole+5FU) is 0.66). Consistent with literature, obstruction, invasion to serosa and contiguous structures, and increased number of lymph nodes involved are associated with higher mortality [125, 126]. However, our spline-based AFT model suggests also that initiating the treatment within 1-3 weeks after the surgery may be associated with 30% longer survival time, although the confidence interval is wide (Figure 4-7). This finding, not reported in the previous analyses of the same data [125, 126], suggests that starting treatment early may maximize the therapeutic benefits. The estimated 5 years survival rates, at mean age and reference values of all binary variables, are 82%, 84% and 90%, respectively, for the observation only, Levamisole and Levamisole+5FU groups. The corresponding survival curve estimates are shown in A-10 in the Appendix.

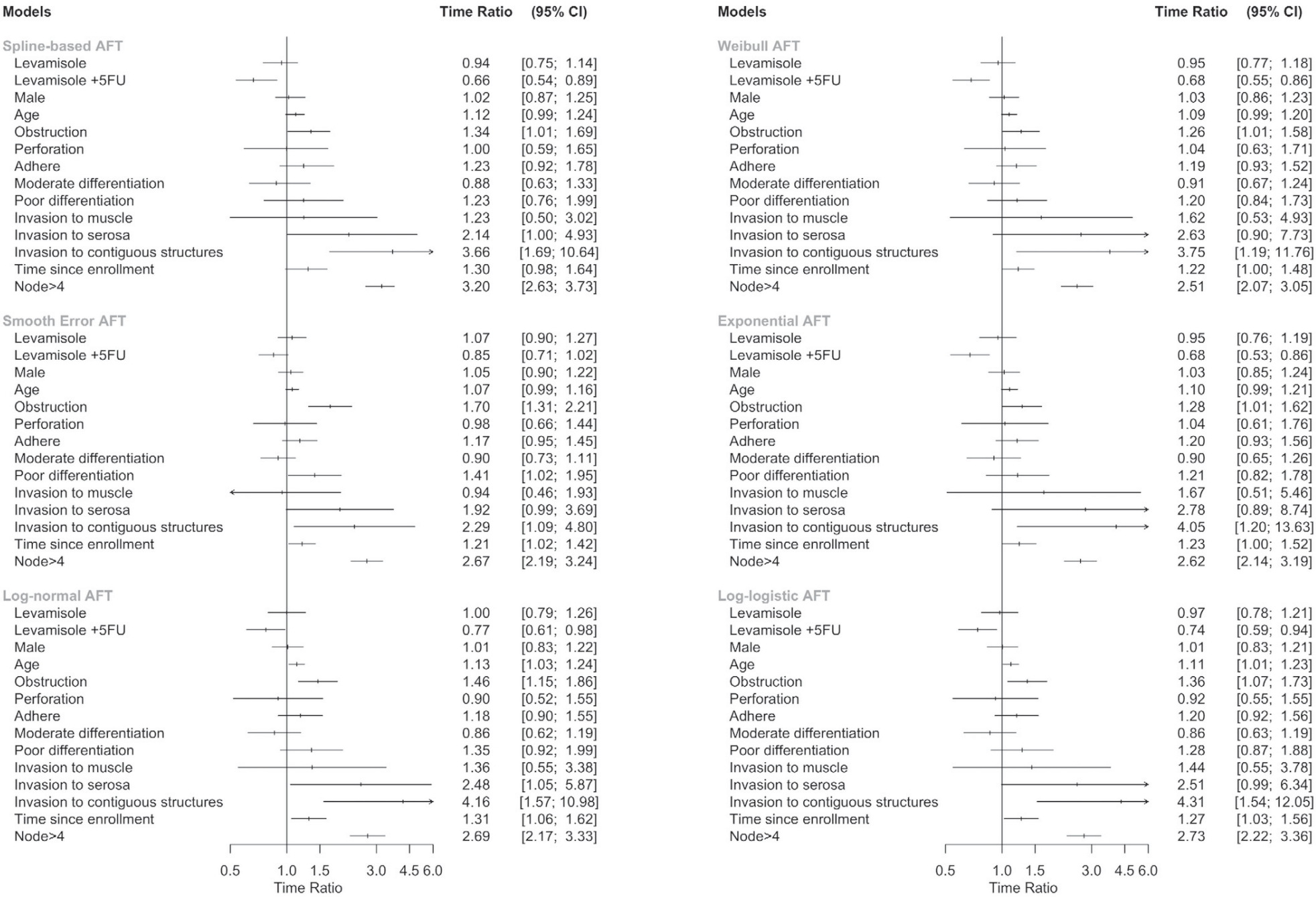


Figure 4-7: Estimated covariate effects in time ratio from alternative AFT models in the colon cancer study.

4.7 Discussion

We proposed a flexible spline-based AFT model, that avoids *a priori* parametric assumptions about the event time distribution. Comprehensive simulations, under different assumptions about the shape of the baseline hazard, indicated that the proposed model yields unbiased estimates of baseline hazard, covariate effects and survival functions, conditional on covariates. Simulations allowed us also to compare the performance of our model with both the conventional parametric AFT models and the semiparametric approach developed by Komárek et al [26].

An accurate assessment of both (i) relative risks, associated with patient characteristics and treatments, and (ii) absolute risks i.e. the probability of survival for patients with different characteristics, are essential for disease prognosis and treatment decisions. One reason why, in clinical and epidemiological applications, AFT models are used considerably less frequently than the Cox PH model, may be related to the difficulties in specifying the distribution of the event times. Indeed, our simulation results confirm that misspecification of this distribution in the AFT analyses may produce biased estimates of both the baseline hazard and survival curves, conditional on covariates. In some AFT-based analyses, this issue was addressed by using regression diagnostics based on comparing the AIC's of alternative parametric models and/or residual plots [127, 128, 129, 130]. However, in complex real-life studies, the baseline hazard may not follow any of the common parametric distributions, implying that more flexible AFT models may be necessary to avoid biased estimates and inaccurate conclusions.

In the last three decades, several flexible extensions of the Cox PH model were proposed to allow assumption-free modeling of both relative risks [40, 81, 101] and absolute risks [41, 103]. In contrast, we found only a few flexible extensions of the AFT model. Whereas some authors used splines to model non-linear effects of continuous covariates on log event time [27, 42, 43], or to relax the variance homogeneity assumption [131], they did not describe flexible modeling of the baseline hazard. Indeed, to our knowledge, Komárek et al.'s smoothed

error semiparametric model [26] may be the only frequentist AFT model, that can be easily implemented to estimate the baseline hazard without parametric assumptions. The spline-based AFT model we propose offers a flexible alternative to their elegant model. In our simulations, both the smoothed error [26] and our spline-based AFT models yielded unbiased estimates of the baseline hazard regardless of whether (i) it was more complex than allowed by the common conventional parametric models or (ii) followed a conventional parametric distribution. However, both for simulated and real-life data analyses, our spline-based estimates were more stable than those based on the smoothed error model, with default option for the penalty.

We recognize that a fair comparison of these two flexible AFT models is complicated by their different approaches to control the smoothness of the estimates. The penalization parameter that controls the smoothness of the Komárek et al.'s estimates is chosen through cross-validation, i.e. based on data-dependent criteria [26], although the user may select a different value of the penalization parameter [91]. In contrast, the number of knots ($m = 2$) and degree ($p = 3$) of the splines in our proposed model are specified *a priori*, based on previous experience with regression spline modeling of survival data, that favors relatively parsimonious models [41, 97]. Thus, to further ensure a fair comparison of the two methods, in simulations we re-fitted the smoothed error AFT model with arbitrary variations of the penalization parameter. The results suggest that whereas a strong penalization improves estimation of relatively simple hazards, it produces restrictively inflexible and, thus, seriously biased estimates if the true hazard is complex. On the other hand, a weak penalization increases instability of the estimates. Yet, in most real-life applications, the true shape of the baseline hazard remains unknown. Thus, our limited simulation results suggest that the default cross-validation-based approach seems preferable for most practical implementations of the Komárek et al.'s model. Whereas in sensitivity analyses the users may vary the penalty, this will require further research to justify the choice of the final model. Moreover, further comparisons of the

two flexible AFT models, using both simulated and real-life data, will be necessary, to more systematically assess their relative advantages and weaknesses.

We applied the proposed spline-based AFT model to re-assess survival in stage III colon cancer. The PH assumption was violated for three covariates and rejected by the omnibus test. On the other hand, diagnostics plots did not suggest important deviations from the AFT assumption. Our proposed spline-based AFT model fit these data better than most of the conventional un-penalized AFT parametric models, which were too constrained to recover the non-monotone baseline hazard. Although the log-normal AFT model fit these data almost equally well, the advantage of our flexible spline-based model was that it avoided the need to 'guess' the shape of the baseline hazard. Moreover, our non-monotone spline estimate was quite smooth and 'regular', in contrast to a wiggly estimate yielded by the Komárek et al.'s smoothed error AFT model.

Some limitations of the proposed model, and our simulations, should be recognized. Firstly, the hyper-parameters: the degree of the splines, as well as the number and the location of interior knots, are chosen *a priori*. Further research may assess the potential benefits of considering alternative hyperparameter values and selecting their values *a posteriori*, based on goodness-of-fit or cross-validation, similar to the approach of Komárek et al [26]. Secondly, because we rely on a two-step alternating conditional estimation (ACE) of (i) the regression coefficients and (ii) the spline coefficients that define the baseline hazard, the covariance matrix of all estimable parameters cannot be approximated using standard large-sample inference. Thus, a non-parametric bootstrap is necessary to correctly quantify the sampling variance of the ACE-based estimates and obtain accurate 95% CI's [33, 41]. Thirdly, given the computational effort required to fit our spline-based model in multivariable analyses, our simulations were limited to a few scenarios, with an arbitrary combination of sample size, censoring levels, covariate effects, but with different event time distributions. We were encouraged to observe that the proposed method performed well across all scenarios considered. Nonetheless, further evaluations, under a wider range of assumptions and parameter values, will be necessary

to assess the robustness of our results and conclusions. Finally, our model relies on the constant time ratio assumption, essential for the AFT modeling, which is as arbitrary as the constant hazard ratio assumption imposed by the PH model. Because the AFT assumption was used to generate data in our simulations, the PH assumption was usually violated, except for exponential or Weibull distributions of event times. Accordingly, our simulations do not permit a fair comparison of the AFT vs. PH models. In real-life applications, researchers should attempt to compare whether the AFT or the PH model seems more appropriate for a given dataset, as illustrated in our colon cancer analysis, where regression diagnostics suggested a better fit of the two flexible AFT models. Yet, the issue will be more complex if different covariates affect the survival according to either the PH or the AFT assumption. Arguably, when the AFT constant time ratio assumption holds for most covariates but a few other behave consistently with the PH assumption, flexible time-dependent extensions of the Cox model could accommodate the resulting non-proportional hazards [41, 132, 133, 134, 135]. However, this would require estimating more parameters, and more complex interpretation, than the AFT model. In contrast, to the best of our knowledge, no currently available AFT model can accommodate time-dependent time ratios. To address this challenge, in future work, we plan to develop an even more flexible version of the AFT model that will allow for time-dependent covariate effects, in order to relax the constant time ratio assumption. Furthermore, more formal tests of the constant time ratio assumption, underlying the AFT model, should be developed and validated.

In summary, we believe that our findings demonstrate both the good performance of our flexible AFT model with spline-based estimation of the baseline hazard and its potential benefits in real-life time-to-event analyses. Therefore, we suggest our model may be considered, together with alternative models, such as the smoothed error method of Komárek et al [26], in future applications, especially when the PH assumption may be questionable and the event time distribution is difficult to specify. We also hope our work will stimulate both wider use of the AFT modeling and further methodological developments.

CHAPTER 5

5.1 Preamble to Manuscript 2

This manuscript further extends the research initiated in the first manuscript of my thesis, by addressing additional methodological challenges encountered in the context of AFT modeling of multivariable survival data. Specifically, it identifies two implicit assumptions imposed in the conventional AFT model and proposes a flexible method for relaxing these assumptions. The spline-based AFT model developed in Manuscript 1 relaxes the parametric assumption for the event time distribution. However, it accepts *a priori* that (i) the time ratio is a constant at any given time point during the follow-up, and (ii) the relationship between the continuous covariates and the log event time is linear. Manuscript 2 further extends the model proposed and validated in Manuscript 1, to simultaneously incorporate time-dependent (TD) covariate effects and non-linear (NL) functional forms of continuous covariates. This methodological challenge, to my knowledge, has not been addressed in the current literature on AFT modeling.

In this manuscript, flexible modeling with un-penalized regression B-splines is used to model both the TD and NL effects, as well as the baseline hazard function. The two-step alternating conditional estimation (ACE) algorithm, employed in the first manuscript, is expanded to three steps to estimate all the parameters in the flexible model developed in the second manuscript. Correspondingly, the statistical program developed for manuscript 1 is extended to implement this new method. Simulation studies are conducted to assess its performance under combinations of various plausible shapes of the TD and NL functions, and to compare

the results to those from conventional parametric AFT models. These methods, together with the spline-based AFT model developed in Manuscript 1, are then applied to re-assess the association between the prognostic factors and mortality after septic shock. The application of the flexible model developed in Manuscript 2, is also included in Manuscript 3, where it is compared to alternative models in real-life analyses of survival in a cancer prognostic study. To reduce overfitting in multivariable settings, I propose to use an iterative backward elimination procedure to select the relevant TD and NL effects into the final flexible AFT model based on the Akaike information criterion (AIC). This strategy is employed in real-life applications in both Manuscripts 2 and 3.

This article is in submitted to *Biometrics*. The publications cited in Manuscript 2 are listed in the reference section at the end of this thesis. Appendix B provides additional information, including the elaboration of the model assumptions, details of the estimation algorithm, as well as additional results of the simulation studies and real-life application.

Manuscript II: Flexible Modeling of Time-dependent and Non-linear covariate effects in Accelerated Failure Time model

Menglan Pang[1], Robert Platt[1,2,3], Tibor Schuster[4], Michal Abrahamowicz[1,3]

[1]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

[2]Department of Pediatrics, McGill University, Montreal, Quebec, Canada

[3]The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

[4]Department of Family Medicine, McGill University, Montreal, Quebec, Canada

Abstract

The accelerated failure time (AFT) model is considered an useful alternative to the Cox proportional hazards (PH) model in survival analysis. However, the validity of conclusions regarding the associations of prognostic factors with event times depends on whether the underlying modeling assumptions are met. Flexible methods for relaxing the PH and linearity assumptions in the Cox model have been extensively studied. In contrast, formal investigation of the corresponding assumptions of constant-over-time time ratio and linearity in the AFT model has been limited. Yet, many prognostic factors may have time-dependent and/or non-linear effects. Furthermore, parametric AFT models require correct specification of the baseline hazard function, which is treated as a nuisance parameter in the Cox PH model, and is rarely known in practice. To address these challenges, we propose a flexible AFT model where un-penalized regression B-splines are used to model (i) the baseline hazard function of arbitrary shape, (ii) the time-dependent covariate effects, and (iii) non-linear effects of continuous covariates. Maximum likelihood estimates of all functions are obtained through an iterative alternating conditional estimation algorithm. The accuracy of the estimated functions in multivariable settings is evaluated by simulation studies. To illustrate potential insights that offered by the proposed model, we apply it to re-assess the effects of prognostic factors on mortality after septic shock.

Key words: Accelerated failure time model, time-dependent effect, non-linear effect, regression splines, simulations

5.2 Introduction

The accelerated failure time (AFT) model provides an alternative to the proportional hazards (PH) model to analyze time-to-event data [6, 7]. Instead of the log hazard ratios estimated in the PH model, in the AFT model the covariate effects are expressed directly on the event time scale

and estimated by the log time ratios. For example, a time ratio of 1.25 (treated vs. control), indicates that the time corresponding to any given survival probability is 25% longer for the treated than the control subjects. In contrast to Cox PH model, parametric AFT models requires specification of the event time distribution [7, 54]. Because in many real-life applications it is difficult to select *a priori* an appropriate parametric distribution, several semiparametric AFT models were proposed [12, 15, 19, 26, 136, 137] to avoid such pre-specification, including our recent spline-based model (under revision at *Statistics in Medicine*).

However, all the aforementioned semiparametric AFT models implicitly accept *a priori* the conventional assumptions regarding the way covariates affect survival. Indeed, just as the Cox PH model implies that effects of covariates on the log hazard are both (i) constant-over-time and (ii) linear, the conventional AFT model also implicitly assumes (i) constant time ratios, for all covariates, and (ii) linear relationships between continuous covariates and the logarithm of event time [3]. (Appendix B.1 provides more details about these two AFT assumptions). Yet, both assumptions may be violated in some real-life settings. Accordingly, several flexible extensions of the Cox model have been proposed to relax the constant hazard ratio (HR) and/or linearity assumptions [39, 40, 101, 138, 139]. Real-life applications of such flexible models reported violations of (i) the PH [45, 140] and (ii) linearity assumptions [141] or even both assumptions, for the same continuous covariate [35, 46].

In contrast, relatively little work has been done on flexible modeling of covariate effects in the AFT framework. Specifically, alternative AFT partial linear models have been proposed to estimate non-linear (NL) effects of a continuous covariate on log event time, through spline smoothing or piecewise linear function [27, 42, 43]. However, these non-linear AFT models either require additional assumptions about the error distribution [27] or seem restricted to the univariate setting [42, 43]. We also found a single reference that relaxes the constant log time ratio assumption, but allows it to be only a linear function βt of the follow-up time t [44]. Furthermore, to the best of our knowledge, no published AFT model permits estimating both TD and NL effects of continuous covariates on the log event time. Yet, simulations suggest that,

under the Cox model framework, the relevant TD and NL effects of all continuous covariates should be simultaneously accounted for to avoid biased estimates and possibly inaccurate conclusions [34, 40].

To address these challenges and offer a more accurate assessment of the role of prognostic factors, we propose a flexible AFT model that incorporates both TD and NL covariate effects while leaving the distribution of the event time free of any parametric assumptions. In section 5.3, we define our model and describe the estimation algorithm. Simulation studies are reported in section 5.4. In section 5.5, we apply our flexible AFT model to illustrate a study of mortality after septic shock, and report TD and/or NL effects of some prognostic factors. The paper concludes with a discussion of our results and their implications.

5.3 Methods

In the conventional AFT model, the natural logarithm of the event time, $\log T$, is modeled as a linear function of the covariate vector $\boldsymbol{X}$ [6, 7]:

$$\log T = -\boldsymbol{\beta}\boldsymbol{X} + W \tag{5.3.1}$$

where W, independent of the covariates, is a random error term, and $\boldsymbol{\beta}$ is the vector of regression parameters representing the logarithms of time ratios, which describe how the covariate values are associated with either accelerated or decelerated event time.

Equivalently, the hazard function in the AFT model can be specified as [7]:

$$\lambda(t|\boldsymbol{X}) = \exp(\boldsymbol{\beta}\boldsymbol{X})\lambda_0\big(\exp(\boldsymbol{\beta}\boldsymbol{X})t\big) \tag{5.3.2}$$

where $\lambda_0(t)$ is the baseline hazard function corresponding to $\boldsymbol{X} = 0$, and $\boldsymbol{\beta}$ in (5.3.2) is the vector of the same constant-over-time log(time ratio) parameters as in equation (5.3.1). AFT

models (5.3.1) and (5.3.2) imply that there is a linear relationship between each continuous covariate and the log event time, and the log time ratios $\boldsymbol{\beta}$ for all covariates are constant during the follow-up. In this article, we extend the AFT model to simultaneously relax both conventional assumptions above and to account for possible TD effects and/or NL effects of continuous variables, in right-censored data.

5.3.1 Joint flexible modeling of NL and TD effects in the AFT model

First, to relax the linearity assumption, the AFT model (5.3.1) can be generalized to:

$$\log T = -\sum_j g_j(X_j) + W \tag{5.3.3}$$

Model (5.3.3) is referred to as the AFT partial linear model [27, 42]. The function $g_j(X_j)$ is a possibly NL transformation of a continuous covariate X_j that estimates how the logarithm of event time changes with increasing covariate value. An equivalent NL extension of the hazard function (5.3.2) is:

$$\lambda(t|\boldsymbol{X}) = \exp\Big(\sum_j g_j(X_j)\Big)\lambda_0\Big(\exp\Big(\sum_j g_j(X_j)\Big)t\Big) \tag{5.3.4}$$

However both models (5.3.3) and (5.3.4), as well as other published flexible versions of the AFT model [26, 27, 42], restrict the covariate effects $g_j(X_j)$ to be constant during the follow-up. To relax this assumption, we propose to further extend the non-linear AFT model (5.3.4), to allow for the covariate effect $g_j(X_j)$ to vary over time, by incorporating a flexible function $\beta_j(t)$ that depends on t. Accordingly, we propose the following flexible extension of the AFT model (5.3.5) that allows for both TD effects of all covariates and NL effects of continuous covariates:

$$\lambda(t|\boldsymbol{X}) = \exp\Big(\sum_j \beta_j(t)g_j(X_j)\Big)\lambda_0\Big(\exp\Big(\sum_j \beta_j(t)g_j(X_j)\Big)t\Big) \tag{5.3.5}$$

This extension involves modifying model (5.3.4) rather than equivalent model (5.3.3) because modeling the hazard function avoids complex constrains that are required in modeling the density function of the event time or the survival function [101]. In the resulting flexible multiplicative AFT model (5.3.5), the effect of a continuous variable X_j on the log hazard, at time t, is modeled as a product of two covariate-specific estimable functions: $\beta_j(t)$ and $g_j(X_j)$. Non-linear effects of X_j are expressed through $g_j(X_j)$. The time-dependent function $\beta_j(t)$, allows for the ratios of event times associated with different covariate values to vary overtime and reflects the dynamic changes in the strength of the covariate effect $g_j(X_j)$ on the log hazard. For a binary covariate, $g_j(X_j) = X_j$, so that only the TD effect $\beta_j(t)$ needs to be considered. The function $\lambda_0\big(\exp\big(\sum_j \beta_j(t)g_j(X_j)\big)t\big)$ describes how covariates alter the baseline hazard $\lambda_0(t)$. Specifically, in model (5.3.5), the covariate affects the hazard function by not only shifting the baseline hazard $\lambda_0(t)$ in the time scale horizontally, as reflected by $\lambda_0\big(\exp\big(\sum_j \beta_j(t)g_j(X_j)\big)t\big)$, but also shifting it vertically by a multiplicative factor $\exp\big(\sum_j \beta_j(t)g_j(X_j)\big)$.

We propose to model $\beta_j(t)$ for each covariate X_j, and $g_j(X_j)$, for each continuous covariate X_j, as well as the unknown baseline log-hazard, using low-dimension un-penalized regression B-splines with degree p and m interior knots:

$$g_j(X_j) = \sum_{l=1}^{L} a_{l,j} A_{l,j}(X_j) \tag{5.3.6}$$

$$\beta_j(t) = \sum_{q=1}^{Q} b_{q,j} B_q(t) \tag{5.3.7}$$

$$\lambda_0\Big(\exp\Big(\sum_j \beta_j(t)g_j(X_j)\Big)t\Big) = \exp\Big\{\sum_{k=1}^{K} \gamma_k S_k\Big(\exp\Big(\sum_j \beta_j(t)g_j(X_j)\Big)t\Big)\Big\} \tag{5.3.8}$$

where $A_{l,j}(X_j)$, $B_q(t)$ and $S_k\Big(\exp\Big(\sum_j \beta_j(t)g_j(X_j)\Big)t\Big)$ are the B-spline basis functions, and $a_{l,j}$, $b_{q,j}$ and γ_k are the spline coefficients to be estimated for $g_j(X_j)$, $\beta_j(t)$ and

$\lambda_0\big(\exp\Big(\sum_j \beta_j(t)g_j(X_j)\Big)t\big)$, respectively. To achieve parsimony in the multivariable flexible AFT model, we consider quadratic splines ($p = 2$) with one interior knot ($m = 1$), implying $L = Q = p + m + 1 = 4$ for estimating $\beta_j(t)$ and $g_j(X_j)$. In the context of flexible extensions of the Cox PH model, such 4 degrees-of-freedom (df) spline estimates have been shown to offer sufficient flexibility to accommodate a variety of NL and TD covariate effects in both simulation and applications [34, 35]. The interior knots for the NL effects are fixed at the median of the sample distributions of respective continuous covariates, whereas the knot for all TD effects is fixed at the median of the distribution of the observed follow-up times. In contrast to the NL and TD effects, to allow for more flexibility in estimating the hazard functions, which may often have more complex shapes than the NL and TD covariate effects, we considered cubic splines (degree $p = 3$) with 2 interior knots ($m = 2$) in (5.3.8). Simulations in our previous work (under revision at *Statistics in Medicine*) have shown that this approach yields accurate approximation of baseline hazard with a variety of shapes. In practice, our method allows other choice regarding both the knot number and their locations in Equations (5.3.6)–(5.3.8), as well as the degree of the splines. In addition, one can specify linear and/or time-invariant effects for some covariates if a simpler model is expected, or impose monotonic functions of time or specific functions of covariates if suggested by substantive prior knowledge.

The full log-likelihood based on model (5.3.5) for right-censored data is derived as:

$$\begin{aligned}
\log L &= \sum_{i=1}^{n} \delta_i \log\big(\lambda(t_i|\boldsymbol{X}_i)\big) + \log(S(t_i|\boldsymbol{X}_i)) \\
&= \sum_{i=1}^{n} \Big[\delta_i\Big(\sum_{j=1}^{J}\beta_j(t)g_j(X_{ij}) + \sum_{k=1}^{K}\gamma_k S_k(\exp\Big(\sum_j \beta_j(t)g_j(X_{ij})\Big)t)\Big) \\
&\quad - \int_0^{t_i} \exp\big(\sum_{j=1}^{J}\beta_j(u)g_j(X_{ij})\big)\exp\big(\sum_{k=1}^{K}\gamma_k S_k(\exp\Big(\sum_j \beta_j(u)g_j(X_{ij})\Big)u)\big)du\Big] \qquad (5.3.9)
\end{aligned}$$

where t_i is the observed follow-up time for subject i, and δ_i is an indicator of whether the subject i experiences the event ($\delta_i = 1$) or is censored ($\delta_i = 0$) at that time.

We can estimate the three sets of spline coefficients $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{\gamma}$ by maximizing the above full log-likelihood in (5.3.9). Substituting $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ into (5.3.6) and (5.3.7), the estimates for the NL effect $\hat{g}_j(X_j)$ and TD effect $\hat{\beta}_j(t)$ can be obtained. Moreover, the hazard and survival functions, conditional on an arbitrary given covariate vector $\boldsymbol{X}$, can be computed as (5.3.10) and (5.3.11):

$$\hat{\lambda}(t|\boldsymbol{X}) = \exp\big(\sum_{j=1}^{J} \hat{\beta}_j(u)\hat{g}_j(X_j)\big) \exp\big(\sum_{k=1}^{K} \hat{\gamma}_k S_k(\exp\Big(\sum_{j} \hat{\beta}_j(t)\hat{g}_j(X_j)\Big)t)\big) \tag{5.3.10}$$

where $\hat{\beta}_j(t) = \sum_{q=1}^{Q} \hat{b}_{q,j} B_q(t)$ and $\hat{g}_j(X_j) = \sum_{l=1}^{L} a_{l,j} A_{l,j}(X_j)$.

$$\hat{S}(t|\boldsymbol{X}) = \exp\big(-\int_0^t \hat{\lambda}(u|\boldsymbol{X})du\big) \tag{5.3.11}$$

5.3.2 Alternating conditional estimation

Estimating the parameters $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{\gamma}$ simultaneously, by maximizing the complex likelihood function in (5.3.9) would be challenging. First, $\beta_j(t)$ and $g_j(X_j)$ for the same continuous covariate are multiplied by each other in model (5.3.5), inducing *non-identifiability* [40]. Furthermore, whereas both $\boldsymbol{a}$ and $\boldsymbol{b}$ need to be estimated to capture the NL and TD effects, both vectors must be considered as fixed 'known' values when estimating coefficients in $\boldsymbol{\gamma}$ that define the log hazard function. To address these challenges, we rely on an iterative alternating conditional estimation (ACE) algorithm [33]. The algorithm iterates across the three consecutive steps, each involving estimating only one of the above coefficient vectors, conditional on the previous estimates of the two other vectors.

Details of the ACE algorithm are provided in Appendix B.2.1, and Appendix B.2.2 discusses bootstrap-based pointwise 95% confidence bands around the estimates.

5.3.3 Converting $\beta(t)$ to time-dependent time ratio estimates

The time-dependent effect $\beta_j(t)$ in model (5.3.5) describes the potential changes over time in the effect of X_j, on the hazard scale. However, in contrast to the constant β in the conventional AFT model (5.3.2), $\beta(t)$ in our proposed flexible AFT model (5.3.5) does not represent time-dependent changes in the log of the time ratio. Given a specific value $\hat{\beta}_j(t)$ estimated from model (5.3.5), the corresponding time-dependent time ratio vary not only across time t but also depend on the covariate patterns being compared. We define the time-dependent time ratio, $\psi(q)$, as a function of the quantiles of the event time distribution q (or the corresponding survival probabilities $1-q$). In general, to reconstruct the time-dependent time ratio comparing survival for subjects with two specific covariate patterns, one needs the inverse of the survival function in (5.3.11) that involves a complex transformation of $\beta(t)$ and $g(X)$, and does not have a straightforward analytic expression. Thus, we rely on a grid search approach to find the q-quantile of the survival time. For example, for a setting with two continuous covariates that both have TD and NL effects, to reconstruct the time-dependent time ratio comparing two subgroups $X_1 = x_1 + 1$ vs. $X_1 = x_1$ but the same value of $X_2 = x_2$ in both groups, we need three steps. The first step (1) is to calculate

$$\begin{aligned}
\hat{S}(t|x_1+1, x_2) &= \exp\left\{ -\int_0^t \hat{\lambda}(u|x_1+1, x_2)du \right\} \\
&= \exp\left\{ -\int_0^t \exp\left(\hat{\beta}_1(u)\hat{g}_1(x_1+1) + \hat{\beta}_2(u)\hat{g}_2(x_2)\right) \right. \\
&\qquad \left. \exp\Big(\sum_{k=1}^{K} \gamma_k S_k\big(\exp\left(\hat{\beta}_1(u)\hat{g}_1(x_1+1) + \hat{\beta}_2(u)\hat{g}_2(x_2)\right)u\big)\Big)du \right\},
\end{aligned}$$

and

$$\hat{S}(t|x_1, x_2) = \exp\Bigg\{ - \int_0^t \exp\Big(\hat{\beta}_1(u)\hat{g}_1(x) + \hat{\beta}_2(u)\hat{g}_2(x_2)\Big) \exp\Big(\sum_{k=1}^{K} \gamma_k S_k\big(\exp\big(\hat{\beta}_1(u)\hat{g}_1(x) + \hat{\beta}_2(u)\hat{g}_2(x_2)\big)u\big)\Big) du \Bigg\}$$

for discrete times t with extremely small increments. The following two steps are: (2) search, separately, for the times t_{q1} and t_{q0} such that $\hat{S}(t_{q1}|x_1 + 1, x_2) = \hat{S}(t_{q0}|x_1, x_2) = 1 - q$; (3) calculate $\psi(q) = \frac{t_{q0}}{t_{q1}}$. These calculations are repeated across the relevant range of q values and the resulting function $\psi(q)$ is plotted to describe the time-dependent time ratio for this specific contrast (here: increasing X_1 from x_1 to $(x_1 + 1)$). Similar calculations can be performed for any contrast of interest, but the results will vary depending on the values of all covariates in the model.

5.4 Simulation Studies

5.4.1 Simulations design

To evaluate the performance of the proposed model in multivariable AFT analyses, we simulated a hypothetical cohort study that followed $N = 1,000$ subjects until the occurrence of the event or administrative censoring, at 6 years. Event times were generated from the extended AFT model, conditional on a binary variable X_1 and two continuous variables X_2 and X_3:

$$\lambda(t|\boldsymbol{X}) = \exp\Big(\beta_1(t)X_1 + \beta_2(t)g_2(X_2) + \beta_3(t)g_3(X_3)\Big)\times \lambda_0\Big(\exp\Big(\beta_1(t)X_1 + \beta_2(t)g_2(X_2) + \beta_3(t)g_3(X_3)\Big)t\Big) \tag{5.4.1}$$

The baseline hazard $\lambda_0(z) = \frac{\alpha}{\rho}(\frac{z}{\rho})^{\alpha-1}$ followed the Weibull distribution with pre-specified shape (α) and scale (ρ) parameters. Two simulated scenarios differed with respect to selected

covariate effects and shape parameters of the increasing baseline Weibull hazard ($\rho = 3$ vs. $\rho = 6$), with a common scale $\alpha = 1.5$. Both scenarios assumed a TD effect of, binary X_1 and both TD and NL effects of continuous X_2. X_3 had a TD effect but linear effect in scenario 1, but a constant-over-time NL effect in scenario 2. The true TD and NL covariate effects for Scenarios 1 and 2 are shown in Figures 5-1 and 5-2 , respectively. Appendix B.3.1 provides details of data generation.

5.4.2 Simulation results

For both scenarios, three multivariable AFT models were fit to each of the 100 simulated samples: (i) the 'conventional' parametric Weibull AFT model with linear covariate effects; (ii) the 'extended' Weibull AFT model with non-linear effects for X_2 and X_3; and (iii) our proposed flexible AFT model (5.3.5) with all possible TD and NL effects. For both models (ii) and (iii), non-linear effects were estimated with quadratic splines with one interior knot at the median of the follow-up time distribution. Models (i) and (ii) rely on a Weibull baseline hazard, and force constant time ratios for all covariates. In contrast, our model (iii) estimates TD effects for all three covariates, and NL effects for X_2 and X_3.

Estimation of the TD and NL functions

We first assess whether our model could accurately recover the true shapes of the TD and NL effects of each covariate on the hazard. Figures 5-1 and 5-2 compare the TD and NL estimates (gray curves) from the 100 independent samples against the corresponding true effects (black dashed curves), respectively for scenarios 1 and 2. (The estimates are rescaled, as explained in Appendix B.2.3). In scenario 1, the vast majority of the estimates correctly recover all the TD effects, regardless of their different true shapes (Figures 5-1(a,b,d)). The NL estimates for X_2 and X_3 in Figures 5-1(c) and 5-1(e) are also quite accurate. The TD estimates show more variability in the tails of the time axis, where the events are less frequent and regression splines

are less stable [81]. Yet, the mean values of all the TD and NL spline-based estimates (white curves) trace fairly close the corresponding true functions (black dashed curves), indicating lack of under- or over-fit bias, even for the truly linear effect of X_3 in Figure 5-1(e).

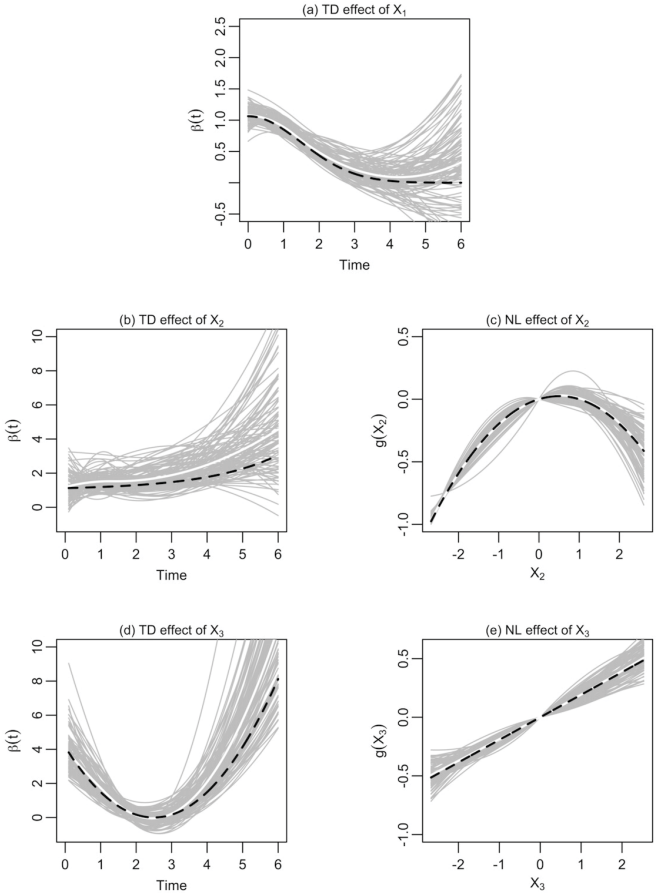


Figure 5-1: Results of the estimated TD and NL effects by the flexible AFT model using 100 samples in simulation scenario 1. The gray curves are the individual estimates from 100 samples, and the pointwise mean is shown by the white curve. The black dashed curve represents the true rescaled NL and TD functions.

Figure 5-2 shows similarly encouraging results for scenario 2. The only exception is that our TD estimates do not recover a decreasing effect of the binary covariate X_1 in the later phase of follow-up, where there are relatively few events (Figure 5-2(a)). All other spline-based TD and NL estimates in Figure 5-2 are reasonably unbiased, including TD estimates of the (truly constant-over-time) effect for X_3 (Figure 5-2(d)).

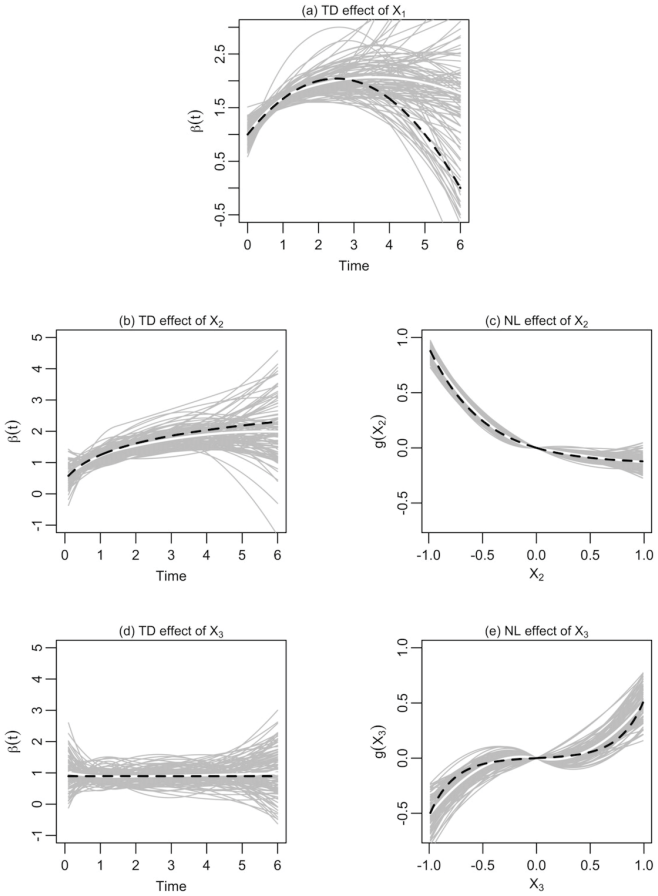


Figure 5-2: Results of the estimated TD and NL effects by the flexible AFT model using 100 samples in simulation scenario 2. The gray curves are the individual estimates from 100 samples, and the pointwise mean is shown by the white curve. The black dashed curve represents the true NL and TD functions.

Appendix B.3.2 provides additional simulation results for scenario 1, with different sample sizes ($N = 650,\ 1,500$) and event rate (40%). Appendix B.3.3 summarizes results from the more constrained AFT models (i) and (ii), that illustrate the impact of misspecification of the covariate effects.

Evaluation of the estimated hazard functions and survival curves

Figure 5-3 compares model-specific estimates of the conditional hazard and survival functions (gray curves) in scenario 1, for the specific covariate vector $X_1 = X_2 = X_3 = 0$, against the corresponding true functions (black dashed curves). Appendix B.3.4 shows similar results for

other selected covariate patterns). For the two Weibull AFT models (i) and (ii), both the hazard (Figures 5-3(a-b)) and the survival function (Figures 5-3(d-e)) estimates are systematically biased, even though event times were generated assuming Weibull baseline distribution. This illustrates the impact of ignoring TD and - for model (i) - NL covariate effects. In contrast, the hazard and survival estimates based on the proposed flexible AFT model (5.3.5), that accounts for the TD and NL covariate effects, do not show any systematic bias (Figures 5-3(c) and 5-3(f)). This further demonstrates the potential advantages of the flexible spline-based modeling in multivariable settings with complex covariate effects. Figure B-11 in the Appendix shows similar results for scenario 2.

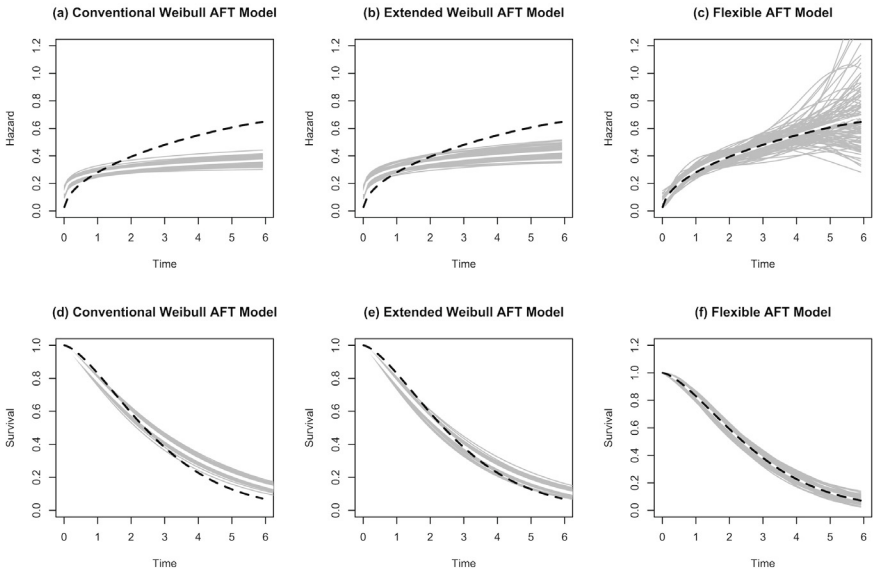


Figure 5-3: Estimated baseline hazard functions (first and third row) and survival curves (second and fourth row) by the three alternative models using 100 samples in simulation scenarios 1. The gray curves are the individual estimates from 100 samples, and the point-wise mean is shown by the white curve. The black dashed curve represents the true baseline hazard and survival functions.

Evaluation of the time-dependent time ratio

Figure 5-4 shows the estimates of the adjusted time-dependent time ratios (gray curves), reconstructed as explained in section 5.3.3, in scenario 1 for specific contrasts between selected

values of each covariate (see Figure 5-4 for details), against the corresponding true time ratios (black dashed curves). The shapes of the time-dependent time ratio estimates for X_1 and X_3 (Figures 5-4(a) and 5-4(c)) generally agree with the corresponding $\beta(t)$ estimates of the TD effects in model (5.3.5) (monotonically decreasing in Figure 5-1(a) and U-shaped in Figure 5-1(d), respectively). The contrast for X_2 has been deliberately chosen so that $g(X_2 = 1) = g(X_2 = 0)$, implying the true time ratio of 1 across the entire follow-up period, and our flexible estimates (gray curves in Figure 5-4(b)) recover well this constant null effect. However, in some situations, the pattern of TD changes in the adjusted time ratios for specific contrasts in the value of a given covariate may (a) substantially diverge from the estimated shape of $\beta(t)$ for the same covariate; and/or (b) vary considerably depending on the values of other covariates. These phenomena are illustrated in Appendix B.5, using selected results of simulation scenarios 1 and 2.

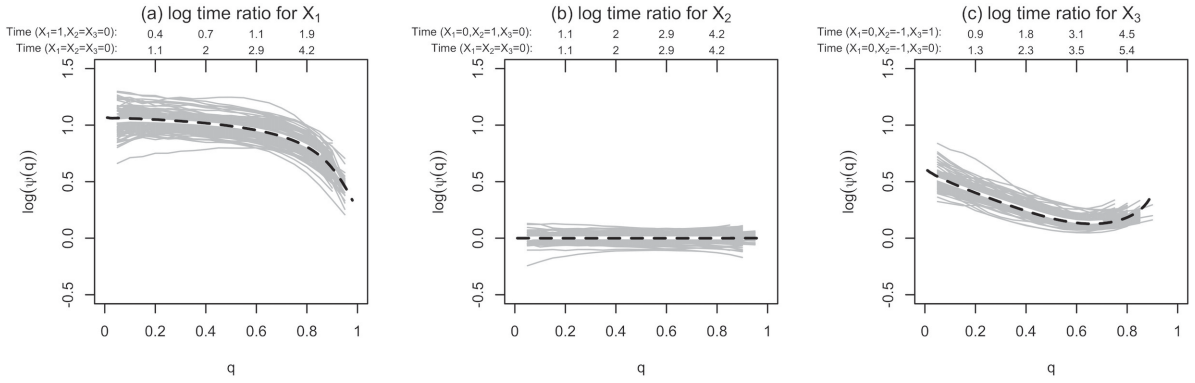


Figure 5-4: Results of the estimated log time ratios by the flexible AFT model using 100 samples in simulation scenario 1, comparing two covariate patterns for each covariate. The two covariate patterns are shown in the labels on the top of each panel, along with the true survival times in both groups corresponding to specific q-quantile of the survival time. The gray curves are the individual estimates from 100 samples, and the pointwise mean is shown by the white curve. The black dashed curve represents the perspective true time ratios.

Comparison of goodness-of-fit

Table 5-1 shows that, as expected, the proposed flexible AFT model fit the simulated data for both scenarios much better than the more constrained AFT models (i) and (ii), with AIC differences between 44 and 125 points. In contrast, in additional simulations, where data were generated from the conventional AFT model (5.3.1), i.e. assuming both constant time ratios and linear effects, our flexible AFT model (iii) yielded AIC worse by 10 and 13 points than the two simpler AFT models (Appendix B.3.6), which correctly suggested the lack of systematic TD and NL effects. Overall, this pattern of results confirms the usefulness of goodness-of-fit comparisons for 'model diagnostics'.

Table 5-1: Comparison of mean akaike information criterion (AIC) in simulation studies from three alternative models

	Flexible AFT model (df=26)	Conventional Weibull AFT model (df=5)	Non-linear Weibull AFT model (df=9)
Scenario 1	2516.16	2641.49	2600.69
Scenario 2	2255.64	2331.04	2299.25

5.5 Real-life Application

5.5.1 Data source

We applied the proposed flexible AFT model to re-analyze 3-month mortality, to investigate the potential TD and NL effects of important prognostic factors, among consecutive adult patients admitted with septic shock to 14 intensive care units (ICUs) in France, between 2009 and 2011 [142, 143]. Time zero corresponded to initiation of vasopressors in response to septic shock, and death of any cause in the next 90 days was the event of interest. Patients alive at 90 days after the septic shock were censored [143]. Details on baseline covariates, measured at admission, were reported elsewhere [142, 143]. Our analyses included 858 patients who had

appropriate antibiotics therapy and complete covariate data. There were 433 (50.5%) deaths during the 1,478 patient-months of follow-up (median duration: 63.5 days).

5.5.2 Flexible AFT analyses

The multivariable analyses included five important baseline prognostic factors, selected *a priori*, based on the published results [142]. Age and the Sepsis-related Organ Failure Assessment (SOFA) score, with higher scores indicating a worse organ dysfunction, were modeled as continuous variables. Binary variables included: immunosuppression (yes vs. no), infection site (urinary tract vs. other), and Knaus score of activity limitations due to chronic health status, dichotomized at normal or moderate (A/B) vs. severe or bedridden (C/D) [142]. Three additional binary covariates, considered for inclusion if they improved the model's fit to data, included: whether or not the germ was identified, infection type (community-acquired vs. nosocomial), and cirrhosis status (yes vs. no).

Appendix B.4.1 describes the 3-stage procedure used to select specific TD and/or NL effects, as well as some of the additional covariates, into the final multivariable flexible AFT model. The pointwise 95% confidence bands for the hazard function, and the TD and NL functions selected in the final model, were estimated through bootstrap, based on 300 resamples.

5.5.3 Alternative models

To assess the impact of the modeling strategy on the results, three additional multivariable AFT models were estimated: (i) the 'conventional' Weibull AFT model, that *a priori* imposed constant time ratios and linear effects of continuous covariates, (ii) the 'extended' Weibull AFT model that allowed for NL effects (see section 5.4 for details), and (iii) the AFT model that imposed linearity but modeled baseline hazard in 5.3.2 with splines, to avoid distributional assumptions (spline-based AFT model developed in manuscript 1). To enhance comparability

of the results, models (i)-(iii) included the same covariates, that were selected into our final flexible AFT model (model (iv)). Notice that models (i)-(iii) could not accommodate potential TD effects and only model (ii) allowed for NL effects.

5.5.4 Results

The spline-based estimate in Figure 5-5(a) indicates that the baseline hazard of all-cause mortality is highest right after septic shock, decreases sharply in the first 10 days, and stabilizes afterwards. This monotonically decreasing hazard could be characterized by a Weibull distribution. For each prognostic factor (row), Table 5-2 compares the effects estimated with the four AFT models (columns). The last column indicates which TD and/or NL effects were selected into our final flexible AFT model 5.3.5. (Table B-2 in the Appendix shows details of model building). For variables for which no TD effects were selected, the estimated constant-over-time time ratios are generally similar across the four models, with shorter survival associated with immunosuppression, non-urinary infections and more severe Knauss scores.

Table 5-2: The estimated covariate effects and akaike information criterion (AIC) values from alternative models in study on mortality after septic shock

Covariates	Conventional Weibull AFT model (df=9)	Non-linear Weibull AFT model (df=13)	Spline-Based AFT model (df=13)	Flexible AFT model†(df=33)
Age	1.05 (1.03, 1.06)	NL	1.01 (0.96, 1.04)	NL+TD
SOFA score	1.51 (1.42, 1.59)	NL	1.32 (1.17, 1.44)	NL+TD
Immunosuppression	2.68 (1.88, 3.81)	2.72 (1.88, 3.81)	2.19 (1.23, 3.11)	2.31 (1.58, 3.25)
Cirrhosis	1.53 (0.90, 2.61)	1.57 (0.98, 2.58)	1.42 (0.84, 2.05)	TD
Knaus score (C/D)	1.79 (1.27, 2.53)	1.70 (1.31, 2.33)	1.75 (1.06, 2.43)	1.51 (1.11, 2.34)
Infection site (Urinary)	0.40 (0.25, 0.63)	0.38 (0.23, 0.67)	0.46 (0.28, 0.80)	0.49 (0.30, 0.69)
Infection type (nosocomial)	1.40 (0.99, 1.98)	1.40 (0.99, 2.00)	1.30 (0.77, 1.61)	TD
AIC	4191.217	4187.541	4195.138	4157.036

Abbreviation: SOFA= Sepsis-related Organ Failure Assessment
†The 95% bootstrap confidence intervals are reported for the spline-based based on 300 bootstrap resamples

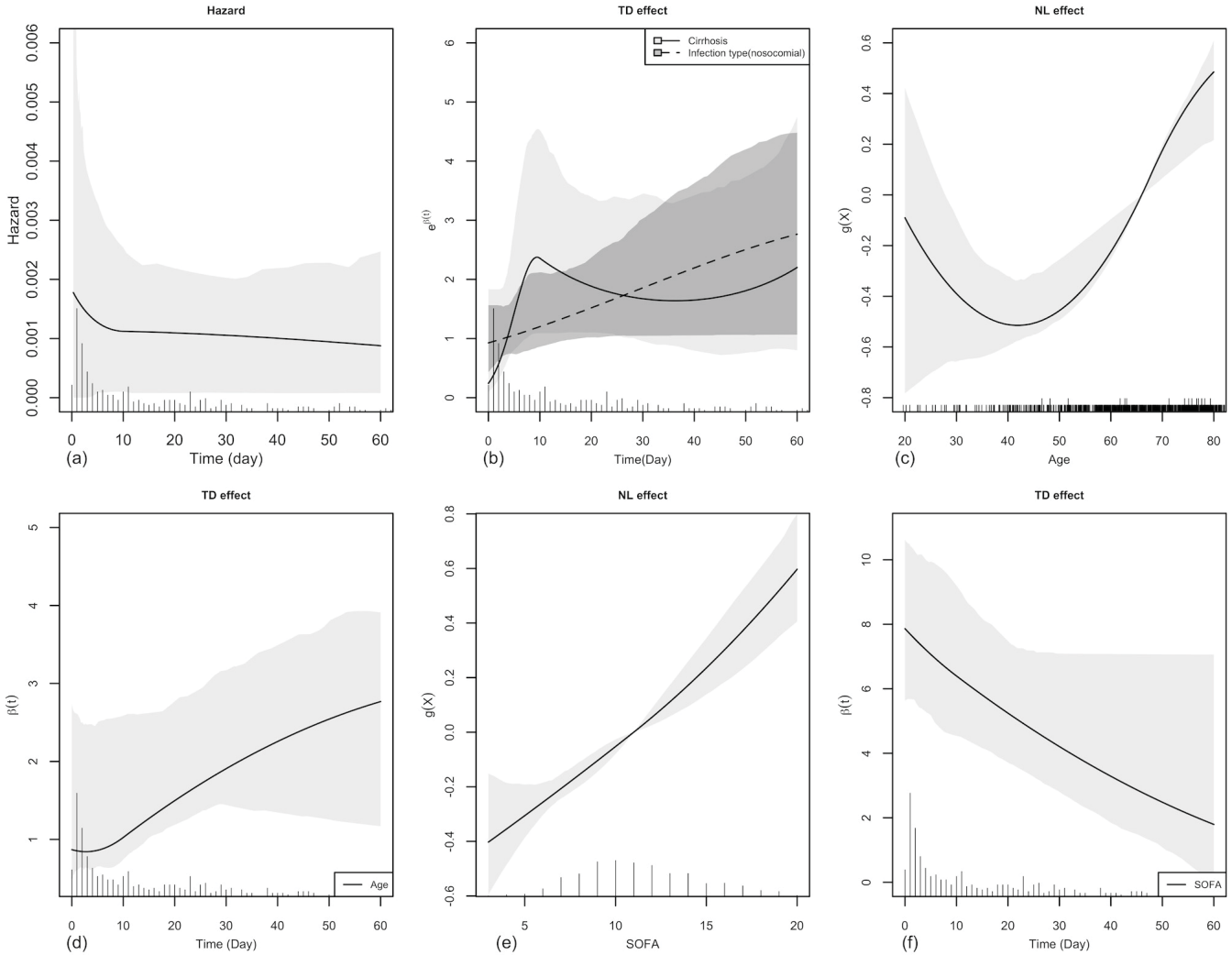


Figure 5-5: Results of the flexible AFT model in study on mortality after septic shock. (a) baseline hazard function (for individuals with all binary covariates equal 0, age equals to the minimal age of 20 years old, and SOFA score equals to the minimal value of 3); (b) TD effects of the cirrhosis status and nosocomial infection; (c) NL effect of age relative to the mean age of 66 years old; (d) TD effect of age; (e) NL effect of SOFA score relative to the mean score of 11; (f) TD effect of SOFA score. Estimates are represented by the black curve, and the shaded grey areas correspond to the 95% pointwise confidence bands, based on 300 bootstrap resamples. The NL effects are constrained to equal 0 at the reference value corresponding to the mean covariate value, and thus the estimates at the reference value show no variation. The empirical distributions of the observed event times (panel a, b, d, f), age (panel c) and SOFA score (panel e) are shown by rug plots at the bottoms of the respective graphs.

Yet, the final multivariable flexible AFT model (iv) included some TD and NL effects (last column of Table 5-2). Specifically, TD effects of nosocomial infection and cirrhosis imply violations of the conventional AFT assumption of constant time ratio. TD estimates in Figure 5-5(b) show how the strength of the two relationships vary over the first 60 days after the septic shock. Nosocomial infection is associated with a higher mortality hazard, and this associations becomes gradually stronger with longer follow-up (dashed curve). This trend is reflected in the time ratios in the first column of Table 5-3. For example, for an 'average' patient who has a community-acquired, rather than a nosocomial infection, the estimated time to reach 85% probability of survival (q=0.15 i.e. $S(t)$=0.85) increases by only 7% (time ratio = 1.07) but the time to $S(t)$=0.7 doubles (time ratio = 1.99). On the other hand, the impact of cirrhosis increases rapidly during the first 10 days and then remains relatively constant (solid curve in Figure 5-5(b)). The fact that both TD effects substantially improve the model's deviance (3-df likelihood ratio test(LRT) statistics of 8.93 for nosocomial infection and 22.24 for cirrhosis), imply also rejection of the corresponding null hypotheses of no association with survival [95]. Indeed, the 95% confidence bands around both TD effects exclude 0 for some portions of the follow-up (Figure 5-5(b)). In contrast, in all simpler models (i-iii), that did not allow for TD effects, the constant-over-time effects of both factors were statistically marginally non-significant (Table 5-2) and their long-term impact seemed to be under-estimated, relative to our TD time ratio estimates in Table 5-3.

Our final flexible model included also NL and TD effects for both age and the SOFA score (last column of Table 5-2). The non-monotone J-shaped NL effect of age (Figure 5-5(c)) suggests mortality is lowest at about 45 years and increases for both younger and older patients. The TD effect of age in Figure 5-5(d) indicates that the impact of age gradually become stronger over time. In contrast, increasing SOFA score has an almost linear effect (Figure 5-5(e)), and the NL effect improves only marginally the model's deviance (4-df LRT=9.256, p=0.055). Whereas this association weakens steadily with increasing follow-up (TD effect in Figure 5-5(f)), a higher SOFA at the time of septic shock is associated with a significant

Table 5-3: Time-dependent time ratios estimated for infection, cirrhosis, SOFA score and age for specific quantiles of the survival time, comparing two subgroups with different values in each of them while having SOFA score and age at the population mean in absence of other binary risk factors.

Quantile (q)	Infection (0 vs. 1)	Cirrhosis (0 vs. 1)	SOFA (11 vs. 12)	Age (65 vs. 70)
0.05	0.95	0.51	1.46	1.21
0.10	0.99	0.84	1.55	1.25
0.15	1.07	1.21	1.61	1.29
0.20	1.20	1.54	1.64	1.36
0.25	1.43	1.89	1.74	1.51
0.30	1.99	2.56	2.07	1.94

Abbreviation: SOFA= Sepsis-related Organ Failure Assessment

risk increase even two months later when the 95% pointwise confidence still excludes 0. The decreasing impact of SOFA score implies that the three other AFT models, constrained by the conventional assumption of constant time ratio, may yield inaccurate estimates of this association. (To further illustrate the changes in the estimated effects of age and SOFA over time, Figure B-15 in the Appendix shows their NL effects estimated at different follow-up times. The NL effects of age and SOFA score estimated by the 'extended' Weibull model (ii) are presented in Figure B-16).

The proposed flexible TD/NL AFT model (iv) offers the best fit to the data, with AIC improved by at least 30 points relative to simpler models (i-iii) (bottom of Table 5-2), even after accounting for the additional 20 df's. This highlights the importance of accounting for the effects NL and/or TD effects of selected prognostic factors. (The Cox-Snell residual plots in Figure B-17 in the Appendix also suggest a better fit of our flexible TD/NL model 5.3.5).

5.6 Discussions

We propose a flexible extension of the AFT model that simultaneously incorporates both the non-linear (NL) and time-dependent (TD) effects of continuous variables on the logarithm of hazard function along with TD effects for categorical variables. The NL and TD effects, as well as the hazard function, are modeled using low-dimension un-penalized regression B-splines. Flexible modeling of the NL effects allows for an arbitrary shape of the dose-response curve, describing how the hazard varies with an increasing value of a continuous covariate. On the other hand, the TD effect permits relaxing the assumption that the coefficient associated with a given covariate is constant and describes how the strength of the covariate effect varies over the follow-up interval. To the best of our knowledge, no previous study has simultaneously investigated both TD and NL covariate effects in the AFT model. Our simulation studies show that the proposed spline-based estimates can accurately recover various plausible shapes of both the NL and TD curves. In addition, the survival curves, as well as the associated time-dependent time ratios, conditional on specific covariate vectors, can be accurately estimated, even in the presence of complex relationships between the covariates and the hazard.

Unlike several flexible models for the time-dependent hazard ratio in the extended Cox model [46, 95, 101, 144], to the best of our knowledge, TD covariate effects in terms of time-dependent time ratios have not yet been addressed in the AFT model literature. One reason may be related to a complex relationship of time-dependent time ratio with the inverse survival function, as outlined in section 5.3.3. In the PH framework, the time-dependent hazard ratio for a given covariate at time t is independent of the baseline hazard, or other covariates. In contrast, in the AFT framework, assessing the time-dependent time ratio for covariate X_j at time t requires inversing the survival probability at time t, $S(t|X)$, which depends on both (i) the underlying baseline hazard and (ii) specific values of all variables included in the model. Therefore, to facilitate separating possibly TD covariate effects from the baseline hazard, and to avoid difficulties in modeling survival and/or density functions [101], we have implemented

our flexible AFT model (5.3.5), by extending the classic AFT model (5.3.2), based on the hazard specification, to replace the constant β by the TD function $\beta(t)$. Our time-dependent AFT model (5.3.5) implies that the instantaneous impact of the covariate X_j on the current hazard, at time t, estimated by $\beta_j(t)$, is common to all subjects, regardless of their values of other covariates, which seems clinically plausible in many real-life applications [95, 140, 145]. However, the complex relationship between covariate effects on the hazard and survival in the AFT model, implies also that the pattern of the resulting time-dependent changes in the event time ratio $\psi_j(q)$ for X_j might not be consistent with the TD effect of X_j ($\beta_j(t)$) on the hazard scale, estimated in equation (5.3.5). Indeed, in section 5.3.3, we show how the numerical transformations necessary to convert the $\beta_j(t)$ estimates into time-dependent event time ratios $\psi_j(q)$, for a specific contrast in the values of covariate X_j, may yield quite different results for subjects with different vectors of other covariates. Figure 5-5 and Table 5-3 illustrate such discrepancies, for the estimated TD effect of the baseline SOFA score on mortality after septic shock. Furthermore, the Figure B-18 and B-19 in the Appendix, provide two artificial analytical examples where the behaviors of (i) $\beta_j(t)$ estimated in equation (5.3.5) vs. (ii) the corresponding reconstructed time-dependent time ratio $\psi_j(q)$ differ substantially. These discrepancies reflect the fact that the time ratio at time t, for a specific contrast in values of X_j, depend partly on the survival before time t and, thus, cannot be interpreted as a reflection of the current impact of X_j, on the instantaneous risk.

Flexible smoothing techniques have been previously adopted in AFT partial linear models to relax the linearity assumption. For example, Orbe et al. uses natural splines to model a non-linear function of a continuous covariate using weighted least squares estimation [27]. However, their approach relies on an additional assumption that the expectation of the error distribution is zero, which is avoided in our model (5.3.5). Zou et al. used penalized splines to model the non-parametric smooth NL function using rank estimation [42]; whereas Xue and et al. proposed to approximate $g(X)$ by a piecewise linear function [43]. However, both Zou et al's and Xue et al's methods are restricted to modeling NL effect of only a single continuous

covariate, in the univariate setting. Moreover, all aforementioned models impose *a priori* the constant time ratio assumption [27, 42, 43]. On the other hand, Elsayed et al. proposed a flexible AFT model with time-dependent coefficients, as a special case of the extended linear hazard regression model, but imposed linearity for the effects of continuous covariates [44]. Our proposed flexible AFT model (5.3.5) is more general than these earlier models as it: (a) relaxes both the linearity and the constant time ratio assumptions, simultaneously, (b) is applicable in multivariable settings. In addition, our method permits estimating survival functions, conditional on possibly NL and/or TD covariate effects, for subjects with any specific covariate vector.

The analyses of mortality after septic shock illustrate the ability of our flexible TD/NL estimates to provide new insights into the role of different prognostic factors. The NL estimate for age suggests a non-monotone relationship between age and hazard, whereas the TD estimate indicates that its strength increases over time. One reason for the increasing TD effect of age may be that mortality soon after a septic shock depends mostly on the severity of the patient's initial condition, largely accounted by clinical indicators (e.g., SOFA and Knaus scores) rather than by age. In contrast, among those who survive this critical early period, older patients are more vulnerable and, thus, more likely to die. Furthermore, TD estimates for both nosocomial infection and cirrhosis suggest that their impact on mortality increases with time since the septic shock (Figure 5-5(b)). In contrast, all simpler AFT models, constrained to constant-over-time event time ratios, suggested that both factors are not associated with survival (Table 5-2).

Some limitations of our proposed method have to be recognized. First, the risk of potential overfit bias is common to most flexible models [39, 81, 92]. In all our analyses of the simulated data, we have *a priori* decided to estimate all potential NL and TD effects, regardless of whether such effects were present or absent in the 'true' data-generating model. We were encouraged to observe that for covariates assumed not to have the non-linear or time-dependent effects, most of the TD estimates were close to constant-in-time effects and the NL estimates generally

approximated well a straight line. Thus, overfit bias was not a major issue in our simulations, with at least ~ 250 un-censored events and three covariates. Additional simulations, with different event frequency and different numbers of covariates, may be, however, necessary to further explore this issue.

Furthermore, due to the complexity of the likelihood function and the iterative ACE procedure, the computation time can be quite long, especially when the dataset is large and TD and/or NL effects of several covariates need to be estimated. For example, the average run time for the 100 simulations for scenario 1 was 1.5 hours on computers with Ubuntu operating system with 3.20 GHz Intel Core i7-8700 CPU and 16 GB memory; whereas it took 6.7 hours to run the final model, with 7 covariates, for the septic shock application on a Mac computer with 2.7 GHz Intel Core i5 CPU and 8 GB memory. Despite the computational burden, we have validated our estimates in simulations, in a relatively complex multivariable setting, under different assumptions about 'true' TD and NL effects. With rapid improvement of the computational power, future real-life analyses of a single, even large, multivariable datasets will become increasingly efficient.

Lastly, further work is needed to systematically compare the proposed flexible TD/NL AFT model with the Cox PH model, and its flexible extensions. Our simulations were designed to evaluate the performance of the proposed model under the AFT framework, therefore the data were generated accordingly. However, in many real-life applications, the 'true' data generating model may be more consistent with the PH model. If so, then the assumption of the constant time ratio underlying the conventional AFT model will be violated, leading to biased estimates, unless the baseline hazard has a relatively simple (exponential or Weibull) form. We expect that, in such situations, by allowing the time ratios to vary during the follow-up, our proposed flexible AFT model with TD effects may still reasonably capture the underlying relationships between the covariates and the hazard, but will require more parameters than the Cox PH model, that will estimate a single hazard ratio for each covariate. A reverse situation will occur if the (unknown) data structure is more consistent with the AFT model.

Therefore, further simulation studies comparing the PH and AFT models, along with their flexible extensions, under a broader range of assumptions concerning 'true' data generating mechanisms, are necessary. For the same reasons, in complex real-life studies with multiple covariates, it is possible that neither the PH nor the AFT assumption is fully satisfied for all covariates. Therefore, we recommend implementing alternative modeling strategies and using goodness-of-fit criteria, supplemented by residual diagnostics, to help choose the final model, or alternative models. However, further simulations may be necessary to systematically evaluate the performance of such criteria and diagnostic tools.

Overall, we have shown that our proposed flexible AFT model yielded reasonably accurate estimates of complex effects of covariates in multivariable analyses. It is readily implemented by our R programs, provided in supporting information. Moreover, our real-life application suggests that the flexible AFT model may offer potential new insights into the role of prognostic factors in clinical studies. Lastly, we hope that our work may help to disseminate and encourage use of AFT modeling in time-to-event analyses.

CHAPTER 6

6.1 Preamble to Manuscript 3

This article compares alternative modeling strategies in survival analysis and illustrates the practice usage of the methods proposed in Manuscript 1 and 2 for survival analysis in a multivariable problem. It illustrates the application of the newly developed complex AFT models in a real-life analysis. It is also motivated by the lack of guidance regarding the choice of the appropriate model in multivariable analysis for time-to-event data.

The method developed in Manuscript 1 can be used to accurately estimate the covariate effects, hazard function, and individual survival curves without imposing parametric assumption for the event time distribution, but imposes conventional linearity and constant time ratio assumptions. Furthermore, Manuscript 2 builds on Manuscript 1 to allow for potential non-linear effects and time-dependent time ratios. On the other hand, the extension of the Cox model by Wynant and Abrahamowicz achieves, within the alternative PH framework, similar flexibility to the new AFT model developed in Manuscript 2. Both models impose only very few assumptions on covariate effects and can accommodate the estimation of hazard function of arbitrary shape. Nevertheless, it is unknown which method fits the empirical data better and provides more reliable predictions for survival, given that usually little prior knowledge is available regarding the way the covariates affect survival. Therefore, both methods, together with other conventional AFT and Cox models, are included in the analysis of the same data in this manuscript.

This research illustrates the real-life usage of the proposed flexible AFT models, which complement the dominant Cox model in the analysis of time-to-event data. It highlights the importance of considering alternative analysis approaches in real-life applications with unknown and potentially complex covariate effects. Moreover, the results in this manuscript re-assess how the prognostic factors affect mortality in non-small cell lung cancer and addresses practical issues regarding the comparison of the performance of different models for survival analysis.

This article is in preparation for submission to the *International Journal of Epidemiology*. The publications cited in Manuscript 3 are listed in the reference section at the end of this thesis. Appendix C provides additional results and interpretations of the real-life application.

Manuscript III: Comparison of Alternative Statistical Models for Re-assessing Survival in Advanced Non-small Cell Lung Cancer

Menglan Pang[1], Robert Platt[1,2,3], Bruno Gagnon[4], Michal Abrahamowicz[1,3]

[1]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

[2]Department of Pediatrics, McGill University, Montreal, Quebec, Canada

[3]The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

[4]4Department of Family Medicine and Emergency Medicine, Faculty of Medicine, Université Laval, Quebec, Quebec, Canada

Abstract

Background: Advanced non-small cell lung cancer (NSCLC) has poor prognosis with short survival time. It is very important to understand the impact of the prognostic factors and identify high risk patients based on their characteristics. Yet, the accuracy of such results depends on whether the model used for the analysis is consistent with the data structure. Most previous studies of survival rely on the conventional Cox proportional hazards (PH) model without checking the crucial PH and linearity assumptions. Violation of these assumptions may lead to important bias and misleading conclusions about the covariate effects and survival estimation. Other survival methods, such as the accelerated failure time (AFT) model, provide alternative frameworks to assess the role of the prognostic factors and perform survival prediction, but also rely on specific assumptions.

Methods: We applied alternative survival models, under either PH or AFT frameworks, to re-assess the associations between several prognostic factors and mortality in non-small cell lung cancer (NSCLC) and used their results to predict individual 1-year survival. Multivariable flexible extensions of Cox PH and AFT models were used to account for potential time-dependent and non-linear effects of continuous covariates, and to estimate the hazard functions and survival curves conditional on these effects.

Results: Flexible survival models suggested short-term prognostic values for several biomarkers, measured from blood test taken at the cohort entry, including albumin, lactate dehydrogenase, absolute neutrophil counts, initial chemotherapy regimen, and C-reactive protein (CRP), as well as non-linear relationships of CRP and age with the mortality hazard. The conclusions regarding the covariate effects were similar between flexible PH and AFT models, both of which yielded slightly better fits to the NSCLC data comparing the conventional methods.

Conclusions: Our study reveals important time-dependent effects of several biomarker on NSCLC mortality. The analyses show the robustness of the results from alternative modeling strategies and demonstrate the importance of flexible modeling in survival analyses.
Key words: Accelerated failure time model, flexible modeling, non-small cell lung cancer, model comparison, survival prediction.

6.2 Introduction

Accurate estimates of individual patients' survival probability are important in clinical research. Non-Small Cell Lung Cancer (NSCLC) is the most common type of lung cancer, comprising 75% to 80% of lung cancer diagnoses, with many patients presenting with advanced stage at their initial diagnosis [146]. Advanced NSCLC has a very poor prognosis with a short median survival time of just four months, and most patients die in less than a year [147, 148]. Identifying patients at a higher risk of early death will allow clinicians to make appropriate treatment decisions [4]. Moreover, it is also important to inform patients of their prognosis, as some patients might decide not to receive treatment at all but instead opt for palliative care [149]. Besides patients' demographic characteristics, relevant prognostic factors are often measured at baseline, including clinical variables that reflect disease severity and important biological or inflammatory markers [35].

To ensure accurate assessment of the relationship between prognostic factors and survival, prognostic studies should rely on statistical methods that represent the true complexity of the baseline hazard function and account for how the prognostic factors affect survival [41]. Yet, popular statistical models for survival analysis, such as PH and AFT models, impose important *a priori* assumptions. More specifically, in the conventional Cox model, the proportional hazards (PH) assumption implies a constant hazard ratio (HR) over time for all covariates, as well as a linear relationship between each continuous covariate and the log hazard [1, 28]. Similarly, in conventional AFT models, the linearity assumption is imposed between continuous covariates

and the log event time [27], and the constant time ratio assumption restricts the log event time to be accelerated or decelerated by a constant covariate effect at any time during the follow-up [3, 72]. In real-life clinical studies, it is not an easy task to check whether the corresponding assumptions for each covariate are consistent with the empirical data, especially when there are many prognostic factors. Diagnostic plots, such as log-log curves and QQ plots, are available for checking the Cox PH and AFT constant time ratio assumptions [52]. However, interpretations of the results of such visual assessment tools heavily depend on subjective judgment and are only applicable to univariate settings, with categorical variables. Moreover, to the best of our knowledge, no statistical test is available to check the AFT constant time ratio assumption, while the Schoenfeld residual test and the omnibus test for detecting the violations of the PH hypothesis are sometimes under powered, especially with heavy censoring [62]. Although we could still use these diagnostic tools as supplementary assessments to explore the data, it is difficult to decide how to choose the final model. Furthermore, once violations have been detected, the data analysts face challenges regarding how to re-model the data to accurately represent complex effects of multiple covariates.

Several flexible extensions of the Cox model have been developed to relax the assumptions [39, 41, 101]. For example, Abrahamowicz and Mackenzie et al. developed a flexible Cox model to allow for time-dependent (TD) effect and non-linear (NL) effects [40], and Wynant and Abrahamowicz [41] proposed a more efficient estimation and extended the model to allow full maximum likelihood estimation (MLE) of the baseline hazard:

$$\lambda(t|\boldsymbol{X}) = \exp\big(\sum_{j} \beta_j^* g_j^*(X_j)\big)\lambda_0(t) \tag{6.2.1}$$

Similar models were suggested by e.g. Sauerbrei et al.[39] and Remontet et al [46].

On the other hand, whereas much less work has been done on flexible extensions of the AFT model, Orbe et al. [27] and Zou et al. [42] proposed some methods to incorporate non-linear covariate effects in the AFT framework. To the best of our knowledge, our recent

work (developed in Manuscript 2) is the first model that accounts for both NL and TD effects within the AFT framework:

$$\lambda(t|\boldsymbol{X}) = \exp\big(\sum_j \beta_j g_j(X_j)\big)\lambda_0\big(\exp(\sum_j \beta_j g_j(X_j))t\big) \tag{6.2.2}$$

In the application of both flexible models (6.2.1) and (6.2.2), data-adaptive approaches are applied to select relevant TD and NL effects of particular variables, using either likelihood ratio tests or goodness-of-fit criteria. This analysis strategy aims to account for TD and NL effects only for those covariates where there is empirical evidence of a meaningful violation of the corresponding assumptions, while reducing overfitting when flexible modeling is not necessary for some other covariates. Moreover, the baseline hazard is estimated in both models, which, combined with the estimated covariate effects, facilitates the estimation of individual survival probability based on subjects' characteristics. The fact that both models (6.2.1) and (6.2.2) are estimated using full MLE facilitates a direct comparison of their goodness-of-fit. Notice that this is *not* possible for comparing AFT models with conventional Cox PH model, due to partial MLE estimation of the latter [1].

Most previous prognostic studies in NSCLC relied on the Cox model without testing the crucial PH assumption [150, 151, 152]. Yet, results obtained based on fitting the flexible Cox model have suggested potentially important violations of the PH and/or linearity assumptions for several covariates [34, 35]. On the other hand, the AFT model, that avoids the PH assumption, has not been frequently applied in real-life prognostic studies. Thus, it remains unclear if the AFT model and its flexible extensions may yield new insights regarding the role of specific prognostic factors on mortality and predicting 1-year survival in advanced NSCLC patients.

The overall aim of this paper is to re-assess the effects of prognostic factors on survival in NSCLC patients using alternative statistical models, in both the PH and the AFT frameworks, and compare the corresponding results, as well as the 1-year survival prediction.

6.3 Data Source

The study cohort consisted of all patients who were diagnosed with advanced NSCLC and treated with chemotherapy between April 9, 2002 and September 18, 2008 at the Jewish General Hospital Pulmonary Oncology Clinic (POC) in Montreal, Quebec, Canada [35]. The date of the first chemotherapy treatment was the cohort entry date, denoted by t_0, and death from any cause was the event of interest. The follow-up was terminated on March 15, 2009 or at 3 years after t_0, whichever comes first, when patients still alive were administrative censored. Several baseline characteristics were collected within 3 days before the start of the chemotherapy, including demographic characteristics, clinical data, and several quantitative biomarkers measured from blood tests. These covariates were considered as potential prognostic factors for NSCLC mortality in the previous study of the same data [35]. After excluding missing data on some of these prognostic factors, we included 269 patients in our study. Descriptive statistics for the baseline characteristics were summarized in the original study [35] and presented in Table C-1 in the Appendix. There were 206 (76.6%) deaths during the follow-up (median duration: 8.8 months). No patients were lost to follow-up.

6.4 Statistical Analysis

6.4.1 Statistical models

We considered six alternative modeling strategies in the analysis, including (i) the conventional Cox PH model [1], (ii) the flexible TD/NL Cox PH model (6.2.1) [41], (iii) the conventional Weibull AFT model [7], (iv) the smoothed error AFT model [26], (v) the spline-based AFT model (developed in Manuscript 1), and (vi) the flexible TD/NL AFT model (6.2.2) (developed in Manuscript 2). Model (i) imposed *a priori* constant HRs and linear effects of continuous covariates in the Cox PH model, whereas model (ii) allowed for the estimation of both TD and NL effects for those covariates for which these assumptions were violated.

Model (iii-v) were all AFT models that imposed *a priori* constant time ratios and linearity assumptions while using different approaches for modeling the baseline hazard. In contrast to model (iii), which assumed a Weibull-based hazard function, model (iv) and (v) used Gaussian mixture [26] and regression B-spline smoothing methods, respectively, to approximate the baseline hazard without specifying the event time distribution. On the other hand, model (vi) additionally incorporated time-dependent time ratio and NL effects of continuous variables, by extending the spline-based AFT model (v). All six models included the same set of baseline prognostic factors, including four categorical variables, i.e., sex, stage (IIIA and IIIB without pleural effusion vs. IIIB with pleural effusion and IV), smoking status (ever vs. never), type of first-line chemotherapy regimen (single vs. double agents), and six continuous variables, i.e., age and C-reactive protein (CRP), lactate dehydrogenase (LDH), albumin, absolute neutrophil counts (ANC) and lymphocytes. Logarithm transformation with base 2 was applied to two continuous variables with highly positively skewed distributions: CRP and LDH.

6.4.2 Model building strategy for the flexible extensions of the PH and AFT models

To alleviate the potential overfitting of the flexible extensions of the multivariable PH and AFT models (ii) and (vi), we adopted an iterative backward elimination procedure to achieve parsimonious final models [34]. The goal of the procedure was to exclude those TD and/or NL effects for which there was little empirical evidence. In contrast, all the initially selected covariates were forced in the final model. At the first iteration, an initial full flexible multivariable model was built to estimate all possible TD and NL effects. At each consecutive step, one among the remaining TD effects of all the covariates or NL effects of all the continuous covariates was excluded. In particular, the TD or NL effect resulting in the largest reduction in Akaike information criterion (AIC) [86] was eliminated. This procedure stopped until there was no further improvement in AIC. The final multivariable models included all the 10

covariates listed in section 6.4.1, but only those TD and NL effects which were *not* eliminated in the backward selection procedure. Non-parametric bootstrap [118], based on 300 resamples, was used to obtain the 95% confidence interval (CI) for the estimates from the final flexible Cox PH (ii) and AFT (vi) models.

6.4.3 Assessment of prediction performance

Prediction of the survival probability at 1 year was performed based on all the six alternative methods. The overall predictive performance was compared by Nagelkerke's R^2 [61] and the Brier score [61, 115]. Discrimination was assessed by Uno's area under the curve (AUC) statistics [113], while calibration performance was evaluated by the calibration slope and intercept [61]. Bootstrap resampling with 300 replications was used for internal validation and prediction performance comparison. The difference in the measures between the bootstrap resamples and the original sample were considered to reflect the optimism level of the original predictive model [61]. Subsequently, for each model, the optimism-corrected measures were obtained, for each index, by subtracting the average optimism from the model's apparent performance [61].

6.5 Results

6.5.1 Association between the prognostic factors and NSCLC mortality

For each prognostic factor, Table 6-1 summarizes its effects estimated with the six alternative survival models. The covariate effects are represented by the adjusted HRs for the Cox PH models (i-ii), and by the adjusted time ratios for the AFT models (iii-vi). The second and the last columns of Table 6-1 indicate which TD or NL effects have been selected into the final flexible TD/NL extensions of the Cox PH model (ii) and AFT model (vi), respectively. (Tables C-2 and C-3 in Appendix C show the results of each step of the backward elimination proce-

dure for these two flexible models). For variables for which no TD effects were selected, the estimated covariate effects are generally consistent across the conventional Cox PH model (i) and its flexible extension (ii), indicating that shorter survival is associated with more advanced tumor stage, smoking, single-agent chemotherapy regimen, and lower value of lymphocytes. Similar effects are indicated by the alternative AFT models (iii-vi).

Table 6-1: Adjusted covariate effects estimated with the alternative multivariable survival models

Variables	Model (i)[a] Cox PH	Model (ii)[a] Flexible TD/NL Cox PH	Model (iii)[b] Weibull AFT	Model (iv)[b] Smoothed error AFT	Model (v)[b] Spline-based AFT	Model (vi)[b] Flexible TD/NL AFT
Stage (IIIB+/IV vs. IIIA/IIIB)	1.78 (1.25, 2.55)	1.85 (1.43, 2.75)	1.60 (1.23, 2.06)	1.54 (1.18, 2.02)	1.54 (1.03, 2.04)	1.58 (1.11, 2.25)
Smoking (ever vs. never)	2.13 (1.35, 3.36)	2.06 (1.39, 3.41)	1.83 (1.32, 2.55)	1.75 (1.25, 2.46)	1.81 (1.25, 2.75)	1.8 (1.21, 2.85)
Chemotherapy (double vs. single)	0.62 (0.43, 0.89)	0.61 (0.41, 0.86)	0.70 (0.54, 0.91)	0.64 (0.48, 0.85)	0.60 (0.47, 0.93)	TD
Sex (female vs. male)	0.99 (0.73, 1.33)	1.05 (0.84, 1.29)	1.00 (0.80, 1.24)	1.08 (0.86, 1.37)	1.06 (0.77, 1.28)	1.02 (0.70, 1.56)
$\log_2$CRP (per doubling of CRP values)	1.12 (1.04, 1.21)	NL	1.09 (1.03, 1.15)	1.10 (1.04, 1.18)	1.06 (1.00, 1.13)	TD
Albumin (per $\uparrow$ [d] of 1 gl^{-1})	0.98 (0.94, 1.02)	TD	0.98 (0.95, 1.01)	0.98 (0.95, 1.02)	0.95 (0.91, 0.97)	TD
$\log_2$LDH (per doubling of LDH values)	2.33 (1.85, 2.94)	TD	1.87 (1.59, 2.19)	1.86 (1.53, 2.26)	1.41 (1.15, 1.71)	TD
ANC (per $\uparrow$ of 3.55×10^9 l^{-1})	1.33 (1.27, 1.39)	TD	1.20 (1.17, 1.24)	1.22 (1.08, 1.37)	1.15 (1.03, 1.33)	TD
Lymphocytesocytes (per $\uparrow$ of 1×10^9 l^{-1})	0.78 (0.62, 0.97)	0.76 (0.61, 0.95)	0.83 (0.71, 0.98)	0.80 (0.68, 0.96)	0.77 (0.61, 0.92)	0.82 (0.66, 0.93)
Age	1.00 (0.99, 1.02)	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)	1.00 (0.98, 1.01)	NL
log-likelihood	-694.62	-680.57	-697.78	-692.83	-694.69	-673.11
df	16	27	12	12.99	16	34
AIC	1421.24	1415.14	1419.56	1411.64*	1421.38	1414.22

Abbreviations: CRP=C-reactive protein; LDH=lactate dehydrogenase; ANC=Absolute neutrophil counts; degrees of freedom (*df*); AIC=Akaike information criterion;
[a]Hazard ratio (HR) and 95% confidence interval (95% CI); [b]Time ratio and 95% confidence interval (95% CI); [d]$\uparrow$: increase. *Not comparable to the other methods.

The final flexible TD/NL extensions of the Cox PH model (model (ii)) and AFT model (model (vi)) are consistent in including TD effects for three common prognostic factors: albumin, $\log_2$LDH, and ANC, indicating that the underlying PH and constant time ratio assumptions are both violated. Figure 6-1 shows how the strength of these covariate effects changes during the follow-up time according to the results from both models (upper panels for the flexible TD/NL Cox PH model (ii) and lower panels for the flexible TD/NL AFT model (vi)). (To avoid very unstable estimates, the graphs are truncated at 18 months of follow-up, after which only 21 deaths were observed.) A higher baseline albumin value is associated with a lower mortality hazard at the cohort entry (Cox's TD effect: HR=0.81 for 1 gl^{-1} with 95% CI: (0.71, 0.95), AFT TD effect: $e^{\beta}(t) = 0.88$ with 95% CI: (0.72, 0.93)). This association becomes gradually weaker and becomes practically null after 6 months, according to both flexible models (ii) and (vi) (Figures 6-1(a)) and 6-1(d)). In contrast, the two estimates of the TD effects of $\log_2$LDH diverge after the first year of follow-up. The time-dependent HR estimated from the flexible model (ii) (Figure 6-1(b)) increases rapidly after 1 year, reaching rather implausibly high values after 15-18 months. On the other hand, the TD effect estimated from the flexible TD/NL AFT model (Figure 6-1(e)) decreases gradually until about 12 months, remaining stable afterwards, which seems more clinically plausible. Nevertheless, both models suggest that the baseline value of this biomarker is a very important predictor of early mortality (Cox's TD effect: HR=2.38 with 95% CI: (1.31, 6.59), AFT TD effect: $e^{\beta}(t) = 2.20$ with 95% CI: (1.23, 3.73) at t_0) and its prediction ability gradually diminishes over the first 12 months. Figure 6-1(c) and (f) show the corresponding TD effects for ANC (for an increase of one standard deviation, i.e., $3.55 \times 10^9 \text{l}^{-1}$). Both flexible models suggest that ANC measured at baseline has a lagged effect on the mortality at the beginning of the follow-up. The impact is initially weak, but it increases over the first 6 months and then gradually declines.

In addition, TD effects are identified by the flexible TD/NL AFT model (vi) for two other prognostic factors: chemotherapy regimen and $\log_2$CRP (Figures 6-2 (b) and (c)). The estimate in Figure 6-2 (b) shows that patients who received double-agents chemotherapy regimen at

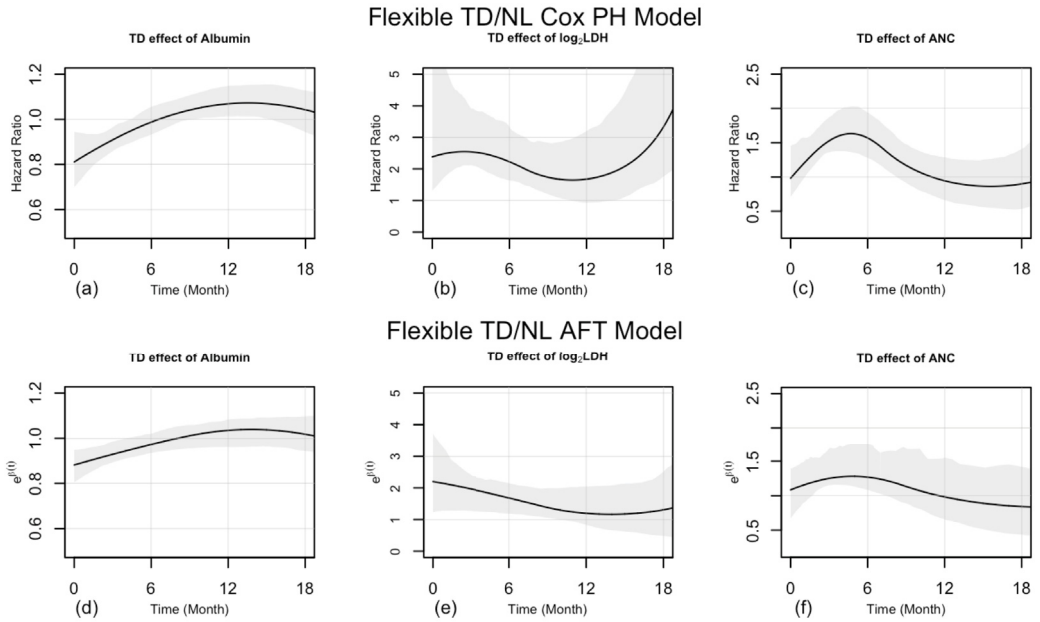


Figure 6-1: The time-dependent effects of albumin, lactate dehydrogenase and absolute neutrophil counts, estimated by the flexible TD/NL Cox PH (upper panels) and flexible TD/NL AFT (lower panels) models.

baseline have considerably lower mortality risk than those who received single-agent regimen. Moreover, a higher value of CRP measured at t_0 is associated with short-term higher mortality risk. Yet, the strengths of these two TD effects also decrease over time, gradually degenerating to nil after about 12 months and 6 months, respectively. In contrast, the TD effects for these two factors were *not* selected in the flexible extension of the Cox PH model (ii). For all the aforementioned factors, simpler models (i, and iii-v) *a priori* constrained their HR's or time ratios to be constant over time, by imposing the corresponding conventional PH and AFT assumptions. The results from these simpler models represent the average effect over time, and thus fail to reflect the gradually attenuated effects of albumin, $\log_2$LDH, chemotherapy regimen, and $\log_2$CRP, as well as the lagged effect of ANC revealed by the flexible TD/NL models (ii) and (vi).

In addition to TD effects of several covariates, the flexible TD/NL models (ii) and (vi) identified non-linear (NL) effects of some continuous variables. Model (ii), that extended

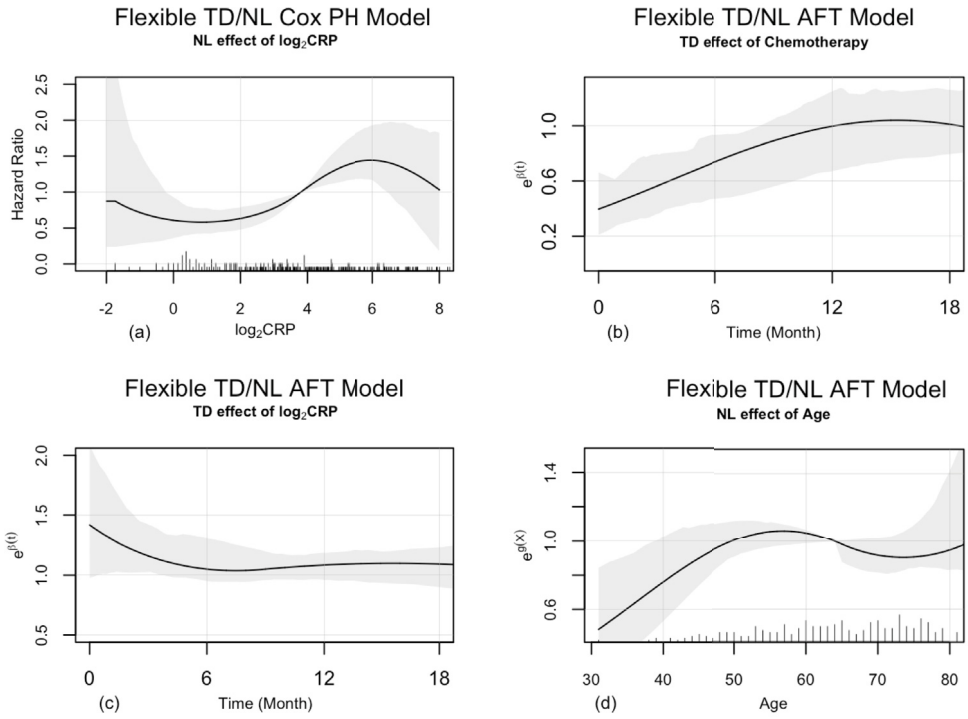


Figure 6-2: The non-linear effect of C-reactive protein (panel (a)) estimated by the flexible TD/NL Cox PH model, and the time-dependent effects of chemotherapy type (panel (b)) and C-reactive protein (panel (c)), as well as the non-linear effect of age (panel (d)), estimated by the flexible TD/NL AFT (lower panels) model.

the Cox PH model, has included NL effect of $\log_2$CRP (Figure 6-2(a)), whereas the flexible TD/NL AFT model (vi) has included NL effect of age (Figure 6-2(d)). (Notice that both NL effects are estimated relative to the mean values of the covariate; therefore, there is no variation at the respective mean (3.8 mgl^{-1} for $\log_2$CRP and 64 years for age). The S-shaped NL effect of $\log_2$CRP, shown in Figure 6-2(a) has been reported in earlier studies [34], and indicates that the mortality hazard increases steeply when $\log_2$CRP at baseline increases from 2 to 6 mgl^{-1}. Some non-monotonicity is exhibited for values outside of this interval but likely reflects numerical instability of the estimate in the regions where data are sparse, as reflected by the wide 95% CI bands in both tails. Figure 6-2(d) implies that mortality is lower for patients diagnosed with NSCLC at younger ages (30-50 years), but remains similar for older patients,

whether 50 or 80 years old at diagnosis. In contrast, all other models (i-v) impose the linearity assumption for age, and, thus, may under-estimate the protective effect of young age.

Some diagnostic plots are presented in Appendix C, providing visual assessment for checking the corresponding PH and AFT constant time ratio assumptions.

6.5.2 Goodness-of-fit of alternative models

Table 6-1 shows that the flexible TD/NL AFT model (vi) fit the data slightly better than the more constrained AFT models (iii and v). The differences in AIC are 5 and 7, after having accounted for 18 and 22 additional degrees of freedom (df's) for the TD and NL effects. Similarly, the flexible TD/NL extension of the Cox PH model (ii) has a slightly better fit than the conventional Cox PH model (i), with AIC reduction of 6.1 points. Interestingly, the two flexible TD/NL models (ii and vi) fit the data comparably under both PH and AFT framework, although the latter included 7 more df's (Table 6-1). Similarly, the AIC values (with the same df's) are almost identical, comparing the conventional Cox PH model (i) and the spline-based AFT model (v). (Details of the calculation of the full log-likelihood for model (i) are provided in Appendix C). The small difference in the AIC goodness-of-fit across the six alternative modeling strategies is also confirmed by the similarly good performance of the corresponding Cox-Snell residual plots (Figure 6-3), in which the two flexible TD/NL models appear to perform slightly better than their more constrained alternatives.

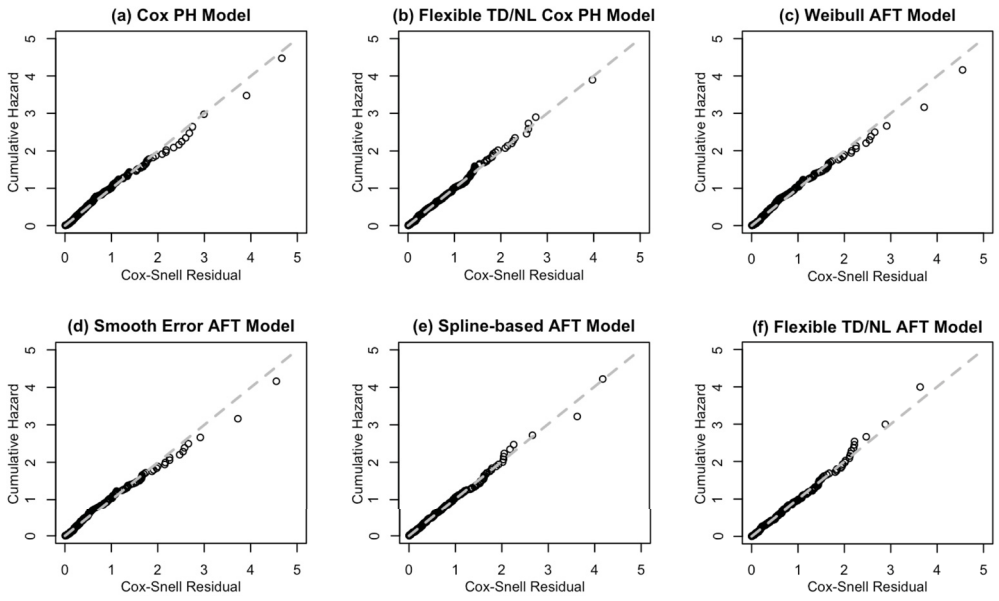


Figure 6-3: Cox-Snell residual plots from alternative survival models

6.5.3 Hazard estimation and survival prediction at one year

Figure 6-4 compares the estimates of the hazard function (left panel) and the corresponding survival curve (right panel) from each of the six alternative survival models for a hypothetical reference group of patients, who are non-smoking men, diagnosed with stage IIIA/IIIB NSCLC, and treated with single-agent chemotherapy regimen at the cohort entry, and mean values of all continuous covariates. For all the models (i) and (iii-v) that do not allow for flexible modeling of the TD and NL effects, the hazard (Figure 6-4(a)) of all-cause mortality steadily increases with increasing time since diagnosis, In contrast, the flexible TD/NL extension of the Cox PH model (ii) shows a temporary decrease in hazard from 6 to 10 months, possibly resulting from the instability of flexible modeling. On the other hand, the flexible TD/NL AFT model (vi) suggests a relatively lower hazard comparing to all other methods, after 1 year. Accordingly, model (vi) suggests higher probability of survival after 1 years since diagnosis (Figure 6-4(b)) than the other models, whereas all methods yield very similar survival curves for the first 6 months of follow-up.

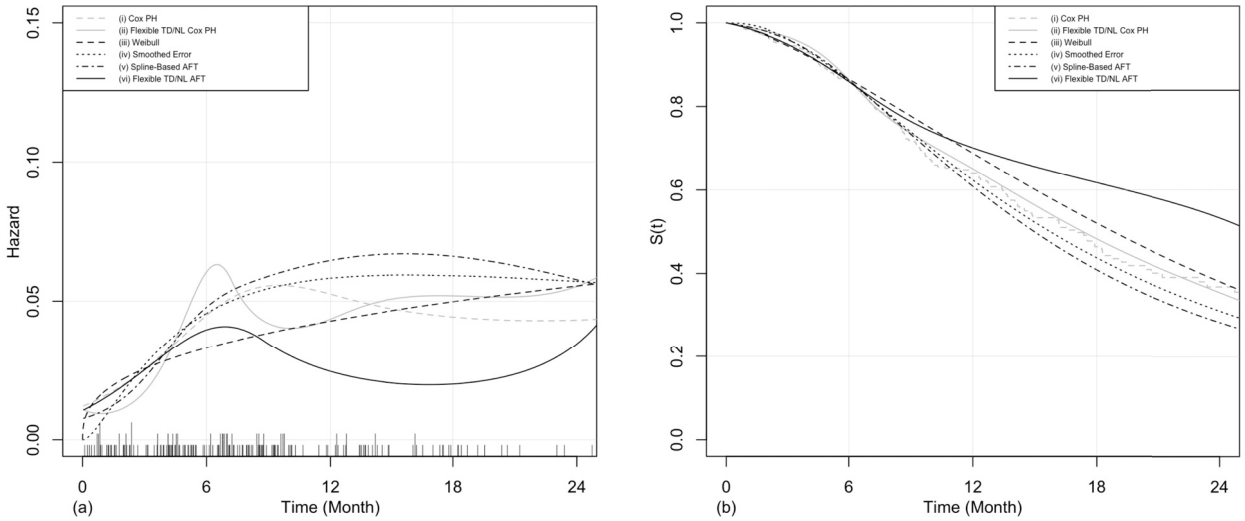


Figure 6-4: The estimated hazard and survival curve by alternative survival models for a reference group with patients who are who are non-smoking men, diagnosed with stage IIIA/IIIB NSCLC, and treated with single-agent chemotherapy regimen at the cohort entry.

Table 6-2 suggests that all the alternative survival model (i-vi) have comparable overall performance on predicting 1-year survival in the NSCLC patients (all Brier scores are approximately 0.2, and Nagelkerke's R^2 ranges from 0.28-0.34). Discrimination is moderately high across all the models, as evidenced by the Uno AUC statistic of approximately 0.74. The flexible TD/NL Cox PH and AFT models (ii and vi) have slightly better calibration than the simpler models (calibration slopes 0.81 and 0.78, and calibration intercepts 0.08 and 0.09, respectively).

Table 6-2: Performance of one-year survival prediction from the alternative multivariable survival models with optimism-corrected measures based on 300 bootstrap validation

	Cox	Flexible TD/NL Cox PH	Weibull	Spline-based AFT	Smoothed error AFT	Flexible TD/NL AFT
Overall model performance						
Brier score	0.208	0.203	0.209	0.210	0.210	0.206
Nagelkerke's R^2	0.340	0.278	0.329	0.340	0.277	0.326
Discrimination						
Uno AUC statistic	0.744	0.750	0.743	0.738	0.743	0.745
Calibration						
Calibration slope	0.775	0.811	0.754	0.745	0.747	0.778
Intercept	0.097	0.084	0.094	0.115	0.105	0.089

6.6 Discussion

Most of the previous studies of NSCLC mortality relied on the conventional Cox PH model and, thus, imposed crucial constant HR and linearity assumptions [153, 154, 155]. We performed comprehensive survival analyses to re-assess mortality in advanced NSCLC using alternative analysis strategies, under both the PH and the AFT frameworks. In particular, the flexible TD/NL extensions of Cox PH and AFT models allow us to simultaneously relax the corresponding constant-over-time effects and linearity assumptions. The results for each of the models, were compared from several angles, including the estimated covariate effects, hazard and survival functions, model's goodness-of-fit, and the prediction of 1-year survival, supplemented by diagnostic plots.

The results from the two flexible models demonstrate their advantages to reveal important TD and NL effects of prognostic factors. Both methods suggest that three baseline biomarkers from blood tests, i.e., lower albumin, and higher LDH and ANC are associated with increased short-term mortality but gradually lose their clinical importance for longer term survival. Short-term unfavorable prognosis associated with low albumin has also been observed in clinical studies [156, 157]. Moreover, similar TD effects of albumin and ANC have been shown in two previous studies that employed the flexible extension of the Cox model in the same data [34, 35], but the effect of LDH has been revealed to remain constantly strong during the entire follow-up period [35]. However, one study [34] has included only four prognostic factors (albumin, $\log_2$CRP, ANC, and smoking), whereas the other one [35] has not adjusted for age and sex, in contrast to our more comprehensive multivariable modeling that revealed important effects of some other prognostic factors. Our flexible TD/NL AFT model also suggests temporary benefit (TD effect) from double-agent regimen at the initiation of the chemotherapy, as well as the protective effect of younger age (NL effect) during the entire follow-up. These flexible TD effects offer more accurate assessments of the prognostic information over time, and the non-linear relationships of both $\log_2$CRP and age with the

hazard may help clinicians to more accurately identify high-risk subgroups. In contrast, all the other conventional methods are unable to capture the dynamic changes of the effects during the follow-up and failed to identify albumin and age as prognostic factors for NSCLC mortality.

On the other hand, many conclusions are robust across all the models we considered. For example, our flexible AFT model (6.2.2) confirmed linear relationships with hazard for 5 of the 6 continuous prognostic factors. Moreover, when both the PH and the constant time ratio assumptions are satisfied, as was the case for many covariates, the estimated effects are generally consistent across the alternative methods. Finally, the estimates of hazard and survival function for a hypothetical reference group, with specific baseline characteristics, are also generally comparable across the alternative models, regardless of different ways to represent covariate effects. These findings, together with the similarity in all models' goodness-of-fit and uniformly good prediction performance, indicate that many conclusions regarding the survival of NSCLC are quite robust. On the other hand, flexible models generated new hypotheses regarding the effects of a few prognostic factors, that should be replicated in further studies, based on independent cohorts.

Our study has some limitations. First, smoothing techniques with regression B-splines are intensively used in the two flexible models (ii) and (vi), which may induce the risk of instability and overfit bias [81, 93] especially given the moderate number of events in our study. Although we have attempted to alleviate this problem by using stepwise backward elimination to reduce the number of parameters in the flexible models, the estimates of the TD effects and the hazard functions exhibit high variability after 18 months of the follow-up, when the number of deaths gradually decreases. Therefore, we only present the results for up to 18 months. Second, it is conceptually challenging to draw final conclusions for the role of each of the prognostic factors in this study, given that the flexible TD/NL extensions of the Cox PH and AFT models fit the data almost identically well but partly diverge regarding selection of TD and/or NL effects of specific covariates. The two modeling frameworks also provide different measures for the relative effects, i.e., HR vs. time ratio, and assume different underlying mechanisms regarding

how the covariates affect the survival [7, 8]. This multiplicity of equally well-fitting models, relying on different assumptions, is common in real-life analysis and has been well recognized [158]. Nevertheless, we argue that it is still better to consider a few alternative analysis strategies and evaluate carefully which model seems more appropriate for the empirical dataset at hand, rather than merely rely on one single model [73]. In the case of similar goodness-of-fit, we demonstrate that our conclusions regarding several prognostic factors are robust across different modeling techniques. Instead of limiting the final interpretation to a single flexible model, we report results from both methods and hope they contribute additional, partly complementary, insights to the current clinical knowledge of NSCLC mortality. Finally, we have not taken into account time-dependent covariates in our analyses due to the unavailability of such data for our cohort [35] and inability of our complex flexible AFT model (6.2.2) to handle them. Time-dependent measures, if available, carry direct information on the time to event [7]. For example, the type and frequency of the chemotherapy regimen may change during the course of the follow-up, and measurements of the biomarkers may be collected regularly over time. These time-dependent measures may be likely associated with the observed event time. Without incorporating such repeated measurements in our analyses, caution is required when interpreting our findings, that are limited to the effects of the baseline values.

In conclusion, our study reflects the analytical and conceptual challenges that investigators would encounter in multivariable analysis of time-to-event data, when considering several alternative survival methods. It illustrates the importance of using flexible survival models to account for the complex effects of prognostic factors that may change over time and/or have potentially non-linear relationships with the hazard. We hope that our results have added new clinical insights into NSCLC mortality and will encourage more widespread applications of flexible modeling of survival data in clinical and epidemiological studies.

CHAPTER 7

Discussion

In epidemiological and medical research with time-to-event outcome, multivariable modeling is essential to assess the covariate effects and to estimate survival probability conditional on individual's characteristics. These estimates, in turn, provide physicians and/or their patients with scientific evidence useful for both prognosis and choice of treatment options. For the past few decades, the Cox proportional hazards (PH) model [1] has dominated the applications of multivariable survival analyses in human health studies [63]. In contrast, the accelerated failure time (AFT) model has been less frequently investigated in theoretical research and applied in real-life analyses. This is perhaps due to the reluctance by investigators to specify a parametric event time distribution, as required by the conventional AFT model [7], in contrast to the Cox PH model that that treats the baseline distribution simply as a nuisance parameter. Indeed, the event time distribution is seldom known in real-life analysis, and its misspecification may lead to biased estimates and misleading conclusions about covariate effects and survival [10, 11]. Several semiparametric AFT approaches have been proposed to tackle this problem, typically deriving ingenious estimation methods that avoid such pre-specification [12, 14, 16, 17, 20, 21]. Nevertheless, most of these methods consider the baseline hazard as a nuisance parameter and, thus, are unable to accommodate estimation of individual survival curves, conditional on covariates. Park and Wei [90] extended the rank procedures [14, 16, 17] to estimate subject-specific survival functions using the Nelson-Aalen estimator of the cumulative hazard function, however, no standard statistical software is available to implement

this estimation. Furthermore, regardless of which model is employed, both the accuracy of model-based estimates and the validity of the inference also depend on whether the underlying assumptions are satisfied. From this perspective, two important implicit assumptions underlie the conventional AFT model: (i) the effect of covariates does not vary over time (constant time ratio assumption) [3], and (ii) the relationship between a continuous covariate and the logarithm of the event time is linear (linearity assumption) [27]. These two AFT assumptions, which are as arbitrary as the PH and linearity of the covariate effects on the log hazard assumptions in the Cox model, have rarely been examined in the statistical literature. Indeed, to the best of my knowledge, no published method allows data analysts to relax both these assumptions simultaneously, in the context of the AFT modeling. Furthermore, in real-life applications, little guideline is available for analysts to decide the analysis strategy considering alternative survival models, including both PH and AFT models or their flexible extensions. Based on these considerations, I have attempted to advance the methodology of AFT modeling to avoid the restrictive assumptions of the conventional model, and to illustrate the practical usefulness of such developments in real-life analyses.

To this end, in this thesis, I have developed and validated novel flexible extensions of the AFT model that do not require pre-specifying the baseline event time distribution and account for potential violations of the constant time ratio and/or linearity assumptions. In other words, the new methods, developed within the AFT framework, aim to (i) provide accurate estimates of the baseline hazard of arbitrary shape, (ii) incorporate the time-dependent (TD) effects and non-linear (NL) effects of continuous covariates simultaneously, and (iii) facilitate the estimation of individual survival curves, conditional on these effects.

The first thesis manuscript focuses on relaxing the parametric assumption about the event time distribution. I have proposed a flexible modeling approach to approximate the baseline hazard of arbitrary shape by using un-penalized low-dimension cubic regression B-spines [92, 93, 94]. To this end, I have derived the close form expression for the full log-likelihood function of the resulting spline-based AFT model. However, estimation of this model's

parameters is not a trivial task, as the covariate effects need to be estimated but, at the same time, must be treated as known 'fixed' values when estimating the spline coefficients that define the baseline hazard. This complication is induced by the way that covariates affect the hazard function in the AFT model. In contrast to the Cox PH model, where the covariate effects can be specified independent of event time distribution (1), covariates (in terms of time ratios) inherently affect the baseline hazard on the time scale in the AFT model (3,12). To address the resulting computational challenge, I have adapted an iterative alternating conditional estimation (ACE) algorithm to estimate parameters of the proposed spline-based AFT model. ACE iterates between the estimation of (i) the covariate effects and (ii) the spline coefficients that define the baseline hazard. In each of the two steps (i) and (ii), only one set of parameters is estimated, conditional on the fixed values of the other parameters, set to their most recent estimates [33]. Simulation studies demonstrated that the proposed spline-based AFT model yielded accurate estimates of covariate effects, baseline hazard, and survival curve, under scenarios with both simple and complex data generating mechanisms.

The second manuscript further advances the methodological developments of Manuscript 1, to relax the constant time ratio and linearity assumptions in the AFT model. The TD and NL functions are incorporated as a product term, and allow for, respectively, the covariate effects to change over time and flexible modeling of possibly non-linear dose-response curves for continuous covariates. Both effects are modeled using quadratic regression B-splines with one interior knot to achieve a parsimonious multivariable flexible TD/NL AFT model, while allowing for baseline hazards of arbitrary shapes. The ACE algorithm employed in Manuscript 1 is expanded to three steps, in order to iteratively estimate three sets of spline coefficients that define the baseline hazard, and the TD and NL functions, respectively. It also avoids a *non-identifiability* problem that occurs in the formulation of the product model, where the estimation of the TD and NL effects are non-separable using a joint likelihood estimation [33, 40]. To validate this new flexible TD/NL AFT model, I have designed a series of novel simulations, assuming complex relationships between the covariates and the hazard,

in multivariable settings. The proposed method and the estimation algorithm yielded overall satisfactory performance. In particular, the flexible estimates yielded reasonably accurate estimates of various, clinically plausible, shapes of both the TD and NL curves, as well as of the individual survival curves, conditional on the multiple covariates and their estimated TD and/or NL effects.

Prior methodological developments on flexible modeling of the AFT model have mostly focused on the challenges addressed in my Manuscript 1. In particular, the extended hazard regression of Etezadi-Amoli and Ciampi [24, 25], that incorporates both the AFT and the PH models, and the smoothed error AFT model of Komárek et al. [26] have shown promising results for covariate effects estimation without the specification of the event time distribution. However, no statistical software is available to implement the extended hazard regression, and the performance of the estimates of baseline hazard and survival functions from both methods have not been investigated in simulation studies. To address this gap in knowledge, in Manuscript 1, I have systematically compared, through a series of simulations, the performance of the smoothed error AFT model [26], my new spline-based AFT model, and several commonly used parametric AFT models. The results indicate that the mis-specified parametric AFT models yield considerable bias in survival estimates, and suggest that, the smoothed error AFT model, with the default option for the penalty of smoothness [26], may provide excessively fluctuating estimates of the baseline hazard. These results help assess the existing methods systematically and support the usage of the proposed spline-based AFT model, which appears to yield uniformly unbiased and numerically stable estimates. On the other hand, a few other existing methods in the current literature have attempted to relax the linearity assumption of the AFT model using flexible smoothing techniques [27, 42, 43]. Yet, most of these methods seem only applicable in the univariate setting. Finally, to the best of my knowledge, only one published paper proposed to relax the constant time ratio assumption, but it restricts the estimated time ratio to be a simple linear function of the follow-up time (25). Moreover, all the aforementioned extensions of the AFT model address only either the linearity or the constant

time ratio assumptions, while imposing *a priori* the other assumption [27, 42, 43, 44]. From this perspective, the flexible TD/NL AFT model developed in Manuscript 2 is more general than these earlier methods and, to my best knowledge, is the first to relax both assumptions simultaneously. The practical importance of such flexible modeling of the covariate effects on time ratio is illustrated by the real-life applications in the analyses of mortality after a septic shock or lung cancer diagnosis that suggest that both TD and/or NL effects are important to accurately describe the role of different prognostic factors (Manuscripts 2 and 3).

The Cox PH model is usually the first choice in the real-life analysis of time-to-event data [63], and the underlying PH and linearity assumptions are often taken for granted [63, 78, 159]. On the other hand, very few real-life studies have explored carefully whether the AFT model could be more appropriate for a given empirical dataset. Given that there is usually little substantive knowledge about the way how covariates affect the survival, it is important for investigators to compare the two models and focus more on interpretation of the results of the model that fits the data at hand better [73]. For example, the results in Manuscript 1 imply that the AFT model is more consistent with the way covariates are associated with survival in colon cancer, where the PH assumption is violated for three prognostic factors. Moreover, in the same analyses, the proposed spline-based AFT model fits the data better than the conventional parametric AFT models. On the other hand, the real-life analyses, based on the proposed flexible AFT model, in Manuscript 2 reveal important TD effects of infection type and Cirrhosis, as well as both TD and NL effects of age and Sepsis-related Organ Failure Assessment (SOFA) score, on mortality after septic shock. Finally, in manuscript 3, the flexible TD/NL AFT model suggests that several biomarkers from blood tests, and the initial type of chemotherapy, have only short-term prognostic utility for predicting mortality in patients diagnosed with non-small cell lung cancer. These new insights offered by the proposed NL/TD flexible AFT model demonstrate its practical usefulness for multivariable real-life analyses of time-to-event data.

On the other hand, the proposed methods and simulation studies have several important limitations. The methods proposed in both Manuscript 1 and 2 involve extensive flexible modeling using un-penalized regression B-splines, which may yield unstable estimates at follow-up times close to the beginning and the end of the study, where only a few events are observed [81]. Furthermore, even if – on average – the estimates are reasonably unbiased, some overfitting bias would likely occur, especially when the number of NL and TD effects included in the final multivariable AFT model is large [93, 160], or in studies with small numbers of events. Thus, to achieve model parsimony and a reasonable bias/variance trade-off, I relied on low-dimension polynomial regression B-splines. Specifically, quadratic splines with one interior knot are used to model the NL and TD curves, whereas cubic splines with two interior knots (that allow more complex shapes [41, 46]) are used to approximate the baseline hazard function. This approach offers sufficient flexibility to recover various clinically plausible functional forms, as indicated by both published simulation studies of spline-based extensions of the PH model [40, 41, 95, 96] and the simulation results presented in Manuscript 1 and 2 of this thesis. However, this *a priori* choice of the model's degrees of freedom should be further assessed, in additional simulations, with an even broader range of true data-generating functions. On the other hand, the existing flexible method by Komárek et al. relies on generalized cross-validation to control the smoothness of the estimates [26]. Given that my dedicated R program allows for arbitrary choice of the number of interior knots and/or the degree of the splines, in future research these hyper-parameters can be selected in a data-adaptive approach, based, for example, on cross-validation. This may be useful for further comparisons with the Komárek et al.'s smoothed error AFT model, that selects the penalty term by cross-validation [26]. One challenge in the development of the proposed methods in this thesis concerns the complexity of the mathematical formulation of the TD and NL effects in the AFT framework. The TD effect is incorporated in the proposed flexible AFT model in Manuscript 2 by replacing the constant covariate effect β with the TD function $\beta(t)$ based on the hazard specification. This approach is adopted to avoid difficulty in modeling

survival or density functions [101], and to facilitate a clinically plausible interpretation of the estimated, possibly time-varying, relationships between the covariate and the instantaneous risk in real-life applications [95, 145]. The resulting TD effect estimate provides a meaningful measure that captures the dynamic change in the strength of the covariate effect during the follow-up. However, it does *not* provide a straightforward estimate for the time-dependent time ratio. Instead, a complex numerical transformation is required to reconstruct the time-dependent time ratio by inversing the resulting survival probability estimates at specific times, as described in section 5.3.3. Moreover, the time-dependent time ratio, by definition, depends on the underlying baseline hazard, as well as on the specific values of all covariates included in the model, and on their estimated, possibly TD and/or NL effects. This conceptual challenge requires more attention when interpreting the results from the flexible AFT model, relative to the conventional alternatives or the flexible extensions of the PH model.

Furthermore, both the analytical and the computational properties of the ACE algorithm, implemented to maximize the complex log-likelihood functions in Manuscript 1 and 2, require careful consideration. It is an iterative conditional procedure; therefore, standard large-sample theory based on the information matrix of the joint log-likelihood cannot correctly quantify the sampling variance of the ACE estimates [41]. Thus, a formal investigation of the asymptotic properties of the proposed estimators is difficult and large-scale, computationally expensive simulations have to be relied upon, as presented in sections 4.4 and 5.4 of my thesis. In addition, the time to convergence can be long, especially with large sample size, and/or large number of covariates with TD and/or NL effects. This computational burden has restricted the number of repetitions to 100 in my simulation studies, and the number of bootstrap resamples to 300 in the real-life applications in each of the three manuscripts. Yet, rapid advances in the computational capacity will improve the efficiency of ACE in future simulation and real-life studies, even with large multivariable datasets.

The main objective of the simulation studies in both Manuscripts 1 and 2 was to assess the proposed flexible extensions of the AFT model. Therefore, the data were generated from the

AFT model or its proposed extensions. Accordingly, it remains unknown how the proposed methods would perform, and how they would compare with alternative survival models if the 'true' data structure is *not* consistent with AFT specification. Further investigation, based on a series of carefully designed simulation studies, is necessary to systematically evaluate and compare alternative regression models for survival analyses, including PH [82], AFT [7, 26], additive hazard [101, 161], and proportional odds (PO) models [103, 104], and, where available, their flexible extensions. In complex real-life studies with many covariates, it is difficult to determine which regression model is most consistent with a given empirical data, and it is plausible that the effects of different covariates may be better accounted for by different models. Therefore, the real-life study presented in each of the three manuscripts includes alternative modeling approaches, including e.g. PH model and/or its extensions in Manuscript 1 and 3, and compares their goodness-of-fit. Nevertheless, the issue of model selection becomes increasingly challenging if we consider an even broader range of survival models, such as other extensions of the PH model [82], flexible parametric PH or PO models [103, 104], or additive hazard models [81, 101, 161]. Another limitation of the methods developed in this thesis is that they do not account for dependent or informative censoring that occurs when the censoring probability or censoring time distribution potentially depends on the event time distribution [7]. The real-life example in each of the three manuscripts assumes independent censoring conditional on the covariates included in the corresponding model. However, additional sensitivity analyses, using inverse probability of censoring weights [162] and non-parametric multiple imputation [163], can be attempted to account for informative censoring, when it may be suspected, based on substantive considerations, data structure and/or study design.

The methods proposed in Manuscript 1 and 2 are mainly developed for clinical or prognostic studies with time-invariant exposure(s) and covariates, measured at the study baseline. They cannot be easily adapted to accommodate time-dependent covariates. In contrast to the PH model, accommodating time-dependent covariates in the AFT framework will demand new

complex re-formulation of the formal model [3, 7]. Furthermore, my research is restricted to time-to-event data with right censoring, although extending to left or interval censoring should be possible by using the corresponding likelihood functions [7]. Thus, interesting directions for future research in this area could involve attempts to generalize further the proposed flexible methods to handle time-dependent data and other types of censoring.

In summary, I have developed, validated and illustrated in real-life applications, new methods to advance the methodology for survival analysis. These methods add to the existing literature on the AFT model to analyze time-to-event data, and their real-life applications may provide new insights in future clinical studies. I hope that my research will stimulate a more widespread use of the AFT model and further methodological developments in this field.

Appendix A: Appendix to Manuscript 1

A.1 Details of the proposed alternating conditional algorithm (ACE) algorithm

Algorithm: The ACE algorithm for estimating the parameters in the proposed spline-based AFT model
Step 1 Initialize $\hat{\boldsymbol{\beta}}^{(0)} = 0$ and $\hat{\boldsymbol{\gamma}}^{(0)} = 1$
For iterations l=1, 2, ...,
Step 2 Set $\hat{\boldsymbol{\eta}}^{(l)} = \hat{\boldsymbol{\beta}}^{(l-1)}$, then estimate $\boldsymbol{\gamma}^{(0)}$ conditional on $\hat{\boldsymbol{\beta}}^{(l-1)}$, and $\hat{\boldsymbol{\eta}}^{(l)}$
Step 3 Estimate $\hat{\boldsymbol{\beta}}^{(l)}$, conditional on $\hat{\boldsymbol{\eta}}^{(l)}$ and $\boldsymbol{\gamma}^{(l)}$
Step 4 Iterate the steps 2 and 3 until the convergence criteria are met

In steps 2 and 3, the Newton-Raphson method is used to obtain, $\boldsymbol{\beta}^{(l)}$ and $\boldsymbol{\gamma}^{(l)}$ using the corresponding derivatives. The first and second derivative with respective to β_r and β_s are given by:

$$\frac{\partial \log L}{\partial \beta_r^{(l)}} = \sum_{i=1}^{n} \Big\{ \delta_i X_{ij} - \int_0^{t_i} \exp\big(\sum_{j=1}^{J} \beta_j^{(l)} X_{ij}\big) \exp\Big(\sum_{k=1}^{K} \hat{\gamma}_k^{(l)} S_k(\hat{w}_i')\Big) X_{ir} du \Big\},$$

and

$$\frac{\partial \log L}{\partial \beta_r^{(l)} \beta_s^{(l)}} = -\sum_{i=1}^{n} \int_0^{t_i} \exp\big(\sum_{j=1}^{J} \beta_j^{(l)} X_{ij}\big) \exp\Big(\sum_{k=1}^{K} \hat{\gamma}_k^{(l)} S_k(\hat{w}_i')\Big) X_{ir} X_{is} du$$

where $\hat{w}_i' = \exp\big(\sum_{j=1}^{J} \hat{\eta}_j^{(l)} X_{ij}\big) u$.

The first and second derivative with respective to γ_r and γ_s are given by:

$$\frac{\partial \log L}{\partial \gamma_r^{(l)}} = \sum_{i=1}^{n} \Big\{ \delta_i S_r(\hat{w}_i) - \int_0^{t_i} \exp\big(\sum_{j=1}^{J} \hat{\beta}_j^{(l-1)} X_{ij}\big) \exp\big(\sum_{k=1}^{K} \hat{\gamma}_k^{(l)} S_k(\hat{w}_i')\big) S_r(\hat{w}_i') du \Big\},$$

where $\hat{w}_i = \exp\big(\sum_{j=1}^{J} \hat{\eta}_j^{(l)} X_{ij}\big) t_i$ and

$$\frac{\partial \log L}{\partial \gamma_r^{(l)} \gamma_s^{(l)}} = -\sum_{i=1}^{n} \int_0^{t_i} \exp\big(\sum_{j=1}^{J} \hat{\beta}_j^{(l-1)} X_{ij}\big) \exp\big(\sum_{k=1}^{K} \hat{\gamma}_k^{(l)} S_k(\hat{w}_i')\big) X_{ir} S_r(\hat{w}_i') S_s(\hat{w}_i') du.$$

A.2 Details of the simulation studies

In simulation A, a complex baseline hazard was defined through a mixture of two Weibull distributions with different shape (α) and scale (ρ) parameters. First, we generated a binary covariate Z with $P(Z = 1) = 0.5$. Then, we assumed (i) $\alpha = 5$ and $\rho = 4$ for subgroup $Z = 0$, versus (ii) $\alpha = 1.5$ and $\rho = 1.2$ for subgroup $Z = 1$. Figure A-1 shows the subgroup-specific baseline hazards and the resulting non-monotonic mixture hazard. The latter generally resembles the all-cause mortality patterns observed in prognostic studies of many cancers, where early high mortality, due to post-surgical complications, is followed by the hazard that first decreases in the next few years and then gradually increases due to aging of the cohort.

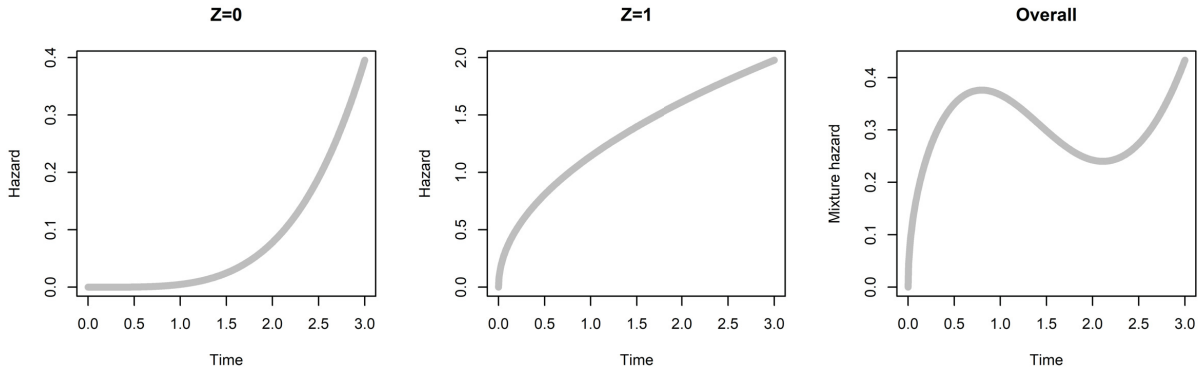


Figure A-1: Two 'component' baseline hazard functions and the resulting overall mixture hazard (rightmost panel) in simulation A.

In contrast to simulation A, simulation B aimed at assessing and comparing the performance of the alternative methods in situations where the flexibility of the smoothed error model and spline-based model were not necessary as the baseline hazard followed one of the classic parametric models. Specifically, individual event times were generated assuming the log-normal baseline distribution with $W \sim \mathcal{N}(0, 1)$. All the other assumptions and the parameter values were identical to simulation A.

For each simulated scenario, the 100 baseline hazards estimated with alternative AFT models, were plotted against the true hazard. Accuracy of the covariate effect estimates was evaluated by relative bias as $\frac{\hat{\beta}_j - \beta_j}{\beta_j} \times 100\%$, empirical standard deviation (SD), and root mean squared error (rMSE). The mean AIC ($-2 \log L + 2df$), across 100 simulated samples, was used to compare the goodness-of-fit of alternative estimation models. (Notice that the number of parameters df varies across models, with $df = 6, 5, 6, 6$, and 10 for Weibull, exponential, log-normal, log-logistic, and the proposed spline-based AFT models, respectively. For the smoothed error distribution model, to reflect the shrinkage of the estimates by penalization, Komárek et al. defined the effective degrees of freedom (df) calculated as the trace of the ratio of two information matrixes of the ordinary likelihood and the penalized likelihood. In our simulations, the resulting AIC values were not comparable to those calculated for the other (un-penalized) AFT models. Therefore, we do not use AIC to compare the fit of the smoothed error vs. the un-penalized AFT models.)

Survival curves for an arbitrarily selected covariate pattern, $X_1 = 1$, $X_2 = X_3 = X_4 = 0$, were estimated based on the results of each model. Specifically, using existing R packages [71], survival curves were estimated based on equation (4.3.4) for parametric AFT models, the Breslow estimator for the Cox model, and the fitted error distribution for the smoothed error AFT model [26]. For the proposed spline-based AFT model (4.3.7), the survival curves were computed with our R program using equation (4.3.11). For each estimation model, the corresponding estimated survival curves from the 100 samples were plotted against the true

survival function. The average relative bias $\frac{\hat{S}(t|X)-S(t|X)}{S(t|X)} \times 100\%$ and the empirical standard deviations were calculated at different time points.

A.3 Additional simulation results from simulation A

Table A-1: The relative bias (%) in covariate effects estimates by parametric AFT models in simulation A based on 500 samples. (Mixture baseline hazard is the true data generating model)

					Weibull				Exponential				Log-normal				Log-logistic			
β_1	β_2	β_3	β_4		X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4
0	0	0	0	n=200	-1.05	0.24	-0.05	-0.66	-1.23	0.29	-0.08	-0.79	-0.8	-0.14	-0.19	-0.88	-1.07	-0.01	-0.16	-0.79
				n=500	-0.8	-0.83	-0.14	0.01	-0.9	-0.95	-0.17	-0.04	-0.83	-0.82	-0.34	-0.27	-0.86	-0.86	-0.23	-0.12
				n=1000	0.36	-0.1	-0.07	-0.01	0.37	-0.12	-0.08	-0.02	0.43	-0.09	-0.11	-0.14	0.41	-0.08	-0.09	-0.08
0.5	-0.5	0.5	0.5	n=200	**7**	**6.7**	2.6	4.5	**21.3**	**20.7**	**8.3**	**9.3**	**10.4**	**11**	3	**5.3**	**5.2**	**5.5**	0.6	3.7
				n=500	**6.6**	**7.2**	4.4	3.7	**19.9**	**20.8**	**10**	**8.2**	**10.2**	**10.6**	**5.9**	4.5	4.9	**5.6**	3.3	2.6
				n=1000	**5.7**	**5.8**	4.2	3.5	**19**	**19.1**	**9.8**	**7.9**	**8.7**	**9.2**	**5.6**	4.7	3.6	3.9	2.9	2.6
0.75	-0.75	0.75	0.75	n=200	3	3.5	1.9	2.3	**13.9**	**14.6**	**6**	**5.5**	**5.3**	**6.5**	2.9	2.7	1.6	2.2	1.1	1.6
				n=500	3	3.3	1.9	2.1	**13.8**	**14.1**	**5.9**	**5.4**	**5.7**	**5.5**	2.9	3	1.5	1.7	1.2	1.7
				n=1000	3	2.6	1.9	1.7	**13.7**	**13.2**	**5.9**	**4.9**	**5.6**	4.9	3.2	2.6	1.5	1	1.4	1.2
1	-1	1	1	n=200	1.7	1.6	1.5	1.5	**10.6**	**10.5**	4.8	4.1	3.8	3.6	2.5	1.9	0.7	0.5	1.2	1.2
				n=500	2.4	1.5	1.3	1.3	**11.2**	**10.1**	4.4	3.7	4.3	3.5	2.3	1.8	1.2	0.3	1	0.9
				n=1000	1.5	1.8	1.3	1.2	**10**	**10.4**	4.3	3.6	3.2	3.7	2.1	2	0	0.6	0.9	1

*The results are presented as bias in the setting where the true covariate effects are null.

Table A-1 shows the relative bias when the true event time is generated from the mixture distribution under different parameter settings for the covariate effects. All the parametric AFT models provide unbiased estimates when the true effects are null. However, they yield biased estimates (relative bias>5%) in at least one of the parameter settings when the true effects are moderate to large. Remarkably, the exponential AFT model always provides large relative bias across all the settings. The standard errors from all the models are comparable.

Estimated survival curve for two alternative covariate patterns in simulation A

The white curve is the pointwise mean of the estimated individual survival curves from 100 simulated samples (gray curves). The true survival function is represented by the black dashed curve.

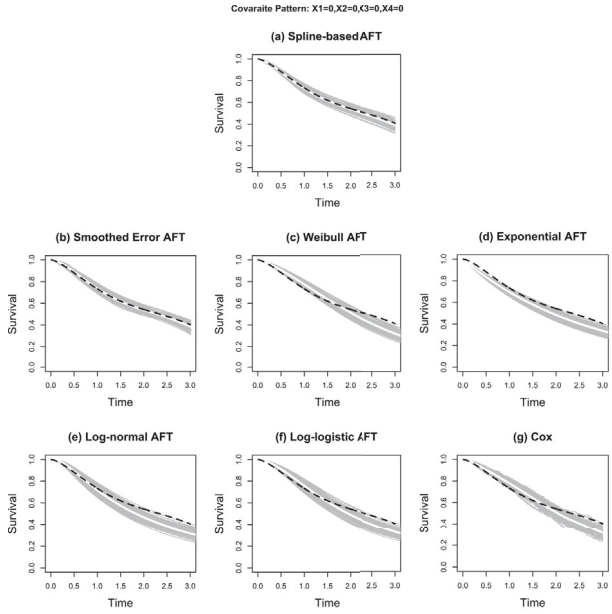


Figure A-2: Comparison of the baseline survival curve estimates, obtained with alternative estimation models (7 panels) in simulation A when the mixture hazard is the true data generating model.

Table A-2: Relative bias (%) and standard error of the estimated baseline survival probabilities from all the alternative model for different time points in simulation A when the mixture hazard is the true data generating model

		Spline-based AFT		Smoothed Error AFT		Weibull AFT		Exponential AFT		Log-normal AFT		Log-logistic AFT		Cox	
	True Survival	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD
t=0.5	0.88	-0.6	0.01	1.1	0.01	2.7	0.01	**-5.5**	0.01	0.9	0.02	3.1	0.02	2.5	0.01
t =1	0.73	-1.2	0.02	0.4	0.02	**5.5**	0.02	**-5.3**	0.02	-2.4	0.03	3	0.03	**6.2**	0.02
t =1.5	0.62	-2.1	0.03	-0.5	0.02	2.9	0.03	**-6.6**	0.02	**-7.5**	0.04	-2	0.04	3.7	0.03
t =2	0.54	**-5.4**	0.03	-1	0.02	**-5.6**	0.03	**-11.1**	0.03	**-14.4**	0.04	**-10.3**	0.04	**-6.2**	0.03
t =2.5	0.48	**10.5**-	0.03	-2	0.02	**-16**	0.03	**-16**	0.03	**-20.3**	0.03	**-17.8**	0.04	**-15.6**	0.03

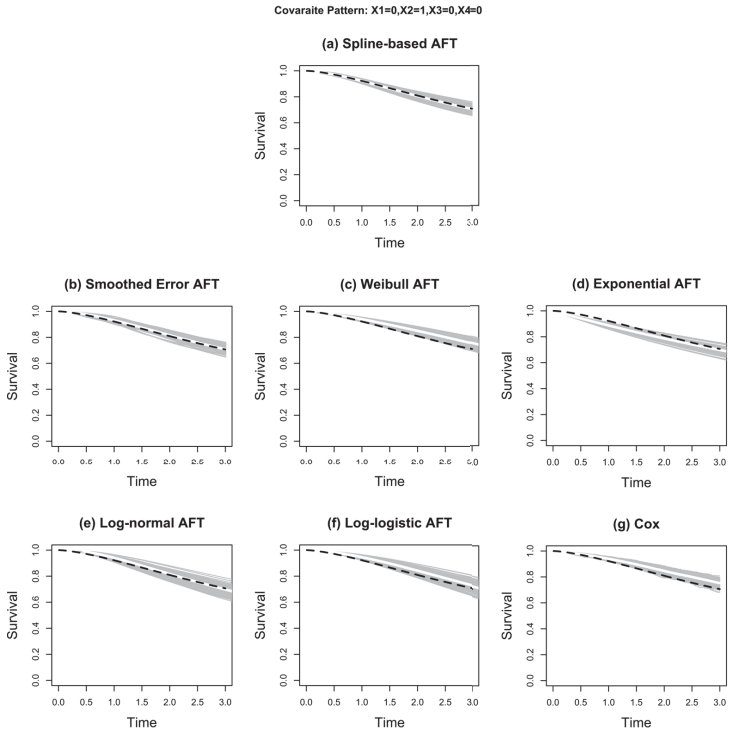


Figure A-3: Comparison of the survival curve estimates, associated with a specific covariate vector ($X_1 = 0$, $X_2 = 1$, $X_3 = X_4 = 0$), obtained with alternative estimation models (7 panels) in simulation A when the mixture hazard is the true data generating model.

Table A-3: Relative bias (%) and standard error of the estimated survival probabilities for covariate pattern X_1=0, X_2=1, X_3=X_4=0 from all the alternative model for different time points in simulation A when the mixture hazard is the true data generating model

		Spline-based AFT		Smoothed Error AFT		Weibull AFT		Exponential AFT		Log-normal AFT		Log-logistic AFT		Cox	
	True Survival	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD
t=0.5	0.97	-0.7	0.01	0.6	0.01	0.5	0	-3	0.01	1.6	0	1.2	0	0.5	0
t =1	0.92	-0.4	0.01	1	0.01	1.8	0.01	-3.9	0.01	1.9	0.01	2.5	0.01	2	0.01
t =1.5	0.87	-0.2	0.02	1.1	0.02	3.4	0.01	-3.7	0.01	1.4	0.02	3.4	0.02	3.5	0.01
t =2	0.81	-0.1	0.02	0.9	0.02	4.9	0.02	-2.9	0.02	0.7	0.03	3.8	0.02	4.6	0.02
t =2.5	0.76	-0.3	0.02	0.6	0.02	**5.9**	0.02	-2.1	0.02	-0.3	0.03	3.5	0.03	**5.9**	0.02

A.4 Additional simulation results from simulation B

Table A-4 shows the relative bias when the true event time is generated from the log-normal distribution under different parameter settings for the covariate effects. All the parametric AFT models provided unbiased estimates when the true effects are null. The mis-specified log-logistic parametric AFT models has good performance across all the settings. On the other hand, the mis-specified Weibull AFT model shows relative noticeable bias, especially when the true effects are small and moderate, and this bias does not diminish with the increase of the sample size. The standard errors from all the models are comparable. Table A-5 summarizes the mis-specified AFT model that produced biased covariate effects estimates (relative bias >5%) under other true event time distributions in at least one of the simulation setting.

Table A-4: The relative bias (%) in covariate effects estimates by parametric AFT models with different event time distribution in simulation B based on 500 samples. (Log-normal is the true data generating model)

True covariate effects					Weibull				Exponential				Log-logistic				Log-normal			
β_1	β_2	β_3	β_4		X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4
0	0	0	0	n=200	0.01	0	0	0	0	0	0	0	0	0.01	0	0	0	0.01	0	0
				n=500	0	0	0	0	0	-0.01	0	0	0	0	0	0	0	0	0	0
				n=1000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.5	-0.5	0.5	0.5	n=200	**-9.0**	**-6.5**	**-5.1**	-2.4	-4.9	-2.4	-3.1	-0.9	-1.0	1.9	-0.4	-0.4	-1.0	1.8	-0.6	-0.4
				n=500	**-6.7**	**-8.1**	-4.6	-3.3	-3.1	-4.5	-2.6	-1.7	0.2	-0.8	0.8	0.1	0.1	-0.7	0.7	0.1
				n=1000	**-6.6**	**-7.6**	-4.6	-4.4	-3.1	-4.1	-2.7	-2.8	0.6	0.0	0.4	-0.3	0.5	0.0	0.5	-0.4
0.75	-0.75	0.75	0.75	n=200	**-5.3**	**-6.8**	-2.6	-2.1	-2.0	-3.6	-1.3	-1.0	0.1	-0.7	0.1	-0.5	0.2	-0.8	0.1	-0.6
				n=500	**-6.3**	**-6.7**	-3.2	-2.5	-3.6	-4.0	-2.0	-1.6	-0.4	-0.8	-0.2	-0.2	-0.4	-0.7	-0.2	-0.2
				n=1000	**-6.0**	**-5.2**	-3.6	-2.4	-3.3	-2.5	-2.4	-1.4	0.1	0.5	-0.3	0.1	0.2	0.7	-0.3	0.2
1	-1	1	1	n=200	-2.9	-4.0	-2.3	-1.3	-0.2	-1.4	-1.3	-0.6	0.8	-0.1	-0.2	-0.4	1.1	0.0	-0.1	-0.3
				n=500	-3.8	-4.3	-2.5	-1.6	-1.6	-2.0	-1.6	-0.9	0.1	0.1	-0.3	0.0	0.4	0.2	-0.2	0.1
				n=1000	-5.0	-4.1	2.1	-1.9	-2.8	-2.0	-1.3	-1.3	-0.9	-0.1	0.1	-0.2	-0.7	0.1	0.1	-0.2

These results indicate that parametric AFT models are robust to certain misspecification in terms of covariate effects estimates for right censor data with random censoring. This is because that except for exponential distribution, the other parametric distributions have two additional parameters (shape or location and scale), which can be estimated in a way to

Table A-5: Misspecification in AFT models with biased covariate effect estimation

True event time distribution	Mis-specified AFT model that yielded biased estimates of covariate effects
Exponential	Log-normal
Weibull	Exponential; Log-normal
Log-normal	Weibull
Log-logistic	Exponential; Weibull
Gamma	Exponential; Log-normal

accommodate the unbiased estimate for the regression coefficients. However, since we rarely know the true event time distribution in practice and not all misspecification have no impact on the covariate effects estimates, it is not recommended to blindly choose a parametric AFT model without strong knowledge about the shape of the hazard. Moreover, these findings were drawn only from the scenarios in which the true event time distribution from one of the classic parametric distribution, and there is no guarantee that mis-specified parametric AFT models would always yield the unbiased estimated under unknown but possibly more complex hazard in practice.

Estimated survival curve for two alternative covariate patterns in simulation B (Data generating model: log-normal)

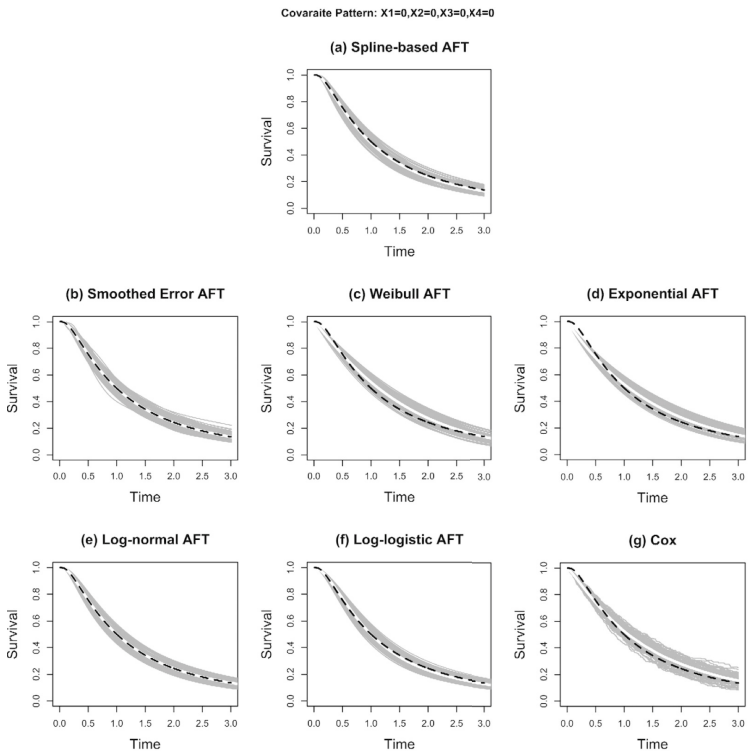


Figure A-4: Comparison of the baseline survival curve estimates, obtained with alternative estimation models (7 panels) in simulation B when the log-normal distribution is the true data generating model.

[Relative bias (%) and standard error of the estimated baseline survival probabilities in simulation B]Relative bias (%) and standard error of the estimated baseline survival probabilities from all the alternative model for different time points in simulation B when the log-normal distribution is the true data generating model

		Spline-based AFT		Smoothed Error AFT		Weibull AFT		Exponential AFT		Log-normal AFT		Log-logistic AFT		Cox	
	True Survival	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD
t=0.5	0.76	0.3	0.03	0.2	0.03	-0.5	0.02	-4.3	0.02	-0.2	0.03	1.4	0.03	-2.1	0.02
t =1	0.50	-0.5	0.03	0.0	0.04	**8.6**	0.03	4.8	0.03	-0.3	0.03	-0.6	0.03	4.6	0.03
t =1.5	0.34	-0.0	0.03	0.1	0.03	**12.6**	0.03	**10.9**	0.03	-0.4	0.03	-4.0	0.03	**10.2**	0.04
t =2	0.24	0.3	0.03	-0.2	0.03	**11.1**	0.03	**12.9**	0.03	-0.4	0.03	**-6.0**	0.03	**13.3**	0.04
t =2.5	0.18	0.5	0.02	-0.3	0.03	**5.3**	0.03	**11.3**	0.03	-0.4	0.02	**-6.4**	0.02	**13.9**	0.04

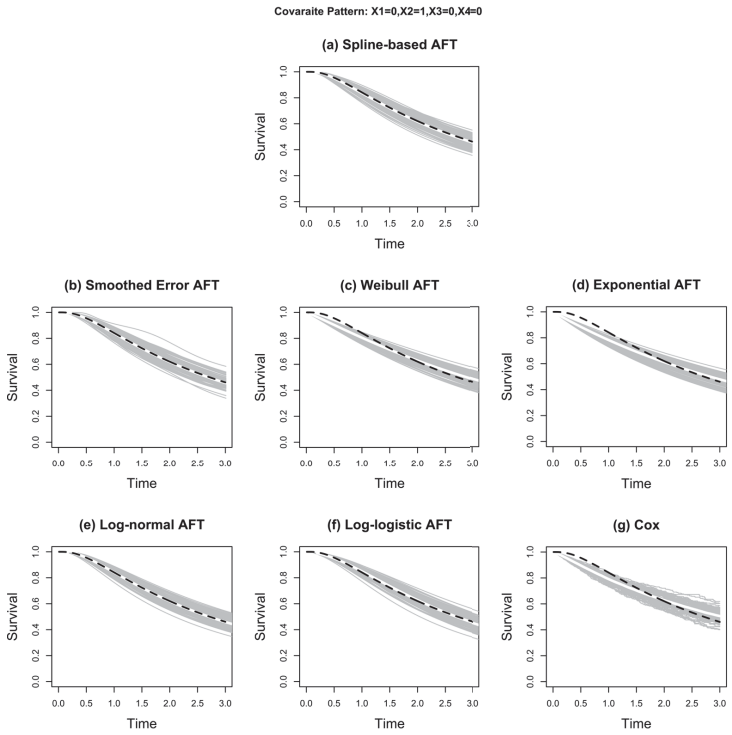


Figure A-5: Comparison of the survival curve estimates, associated with a specific covariate vector ($X_1 = 0$, $X_2 = 1$, $X_3 = X_4 = 0$), obtained with alternative estimation models (7 panels) in simulation B when the log-normal distribution is the true data generating model.

Table A-6: Relative bias (%) and standard error of the estimated survival probabilities for covariate pattern X_1=0, X_2=1, X_3=X_4=0 from all the alternative model for different time points in simulation B when the log-normal distribution is the true data generating model

		Spline-based AFT		Smoothed Error AFT		Weibull AFT		Exponential AFT		Log-normal AFT		Log-logistic AFT		Cox	
	True Survival	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD	Relative Bias (%)	SD
t=0.5	0.95	-0.3	0.01	-0.2	0.01	**-5.4**	0.01	**-7.5**	0.01	-0.2	0.01	-0.7	0.01	**-6.1**	0.02
t =1	0.84	0.5	0.03	-0.4	0.03	-4.5	0.02	**-7.2**	0.02	-0.4	0.02	0.6	0.02	**-5.9**	0.03
t =1.5	0.72	-0.7	0.04	-0.4	0.04	-1.8	0.03	-4.7	0.03	-0.6	0.03	1.1	0.03	-2.8	0.03
t =2	0.62	-1.4	0.04	-0.5	0.04	1.0	0.04	-1.8	0.03	-0.7	0.04	0.5	0.04	1.3	0.04
t =2.5	0.53	-1.7	0.04	-0.8	0.04	3.2	0.04	1.0	0.04	-0.8	0.04	-0.6	0.04	**5.7**	0.04

A.5 Additional Results from the colon cancer study

Table A-7: Distribution of baseline characteristics and pathological variables in the stage III colon cancer study

Continuous Variables (mean (sd))	
Age	59.81 (11.91)
Categorial Variables (N (%))	
Male	460 (51.8)
Treatment	
Observation	305 (34.3)
Levamisole	294 (33.1)
Levamisole+5FU	289 (32.5)
Obstruction	171 (19.3)
Perforation	27 (2.0)
Adherence to nearby organs	128 (14.4)
Histologic differentiation	
Well	90 (10.1)
Moderately well	653 (73.5)
Poor	145 (16.3)
Depth of invasion	
Submucosa	19 (2.14)
Muscle	120 (11.5)
Serosa	730 (82.2)
Contiguous structures	37 (4.2)
Days from surgery to cohort entry=21-35	238 (26.8)
Positive lymph nodes >4	235 (26.5)

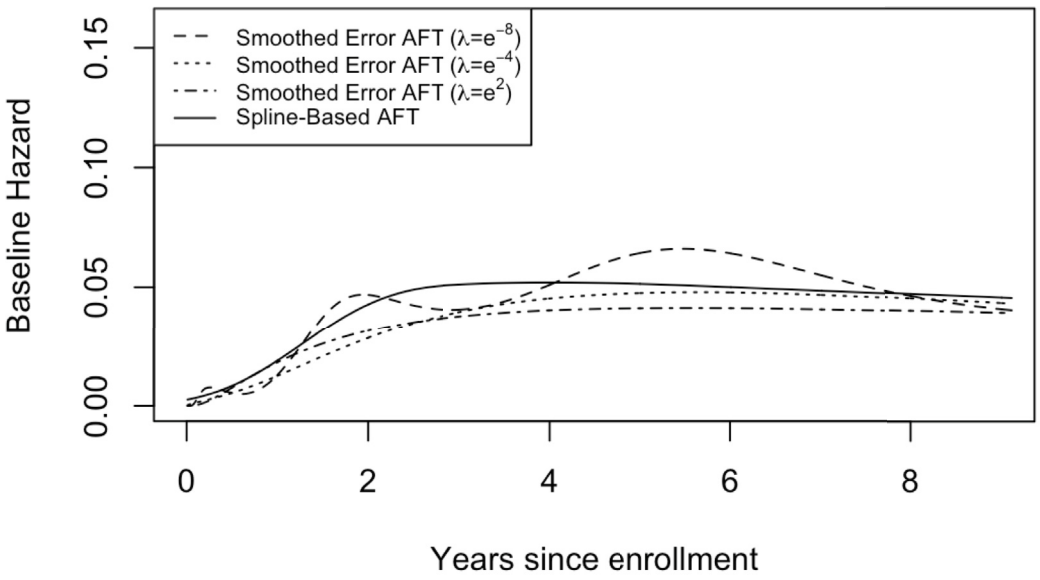


Figure A-6: The baseline hazard of all-cause mortality in the stage III colon cancer, comparing the spline-based AFT model to the smooth error AFT model with increasing penalty.

Table A-8: Models fit of the smoothed error AFT model based on increasing penalty

log(λ)	Log likelihood	Effective degrees of freedom	AIC	Penalized log-likelihood
-8	-1280.18	24.688	2609.74	-1283.93
-4	-1287.67	20.421	2616.17	-1288.34
2	-1293.89	16.204	2620.18	-1294.85

$\lambda = e^{-8}$ is chosen by the default option using cross-validation; number of parameters is 54 for all the three models

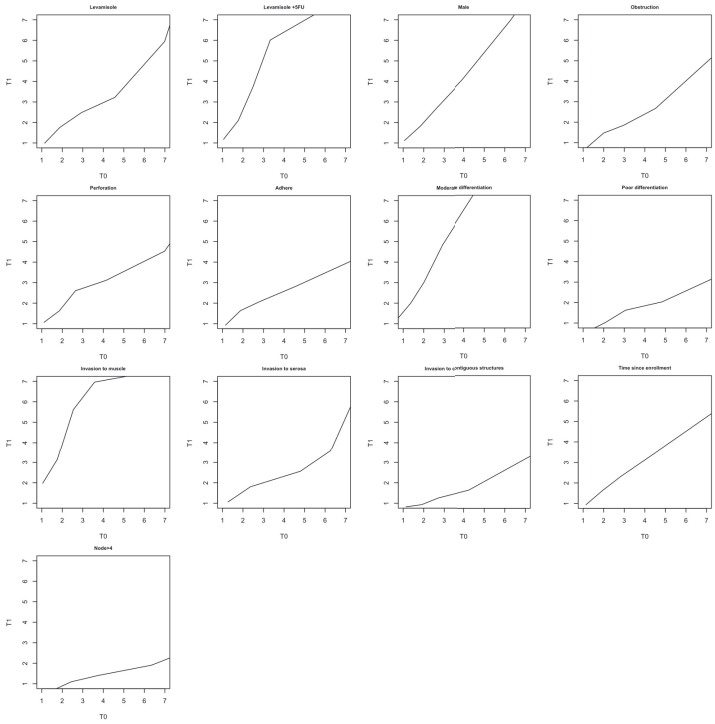


Figure A-7: QQ plots for checking the AFT constant time ratio assumptions for each covariate in the colon cancer study.

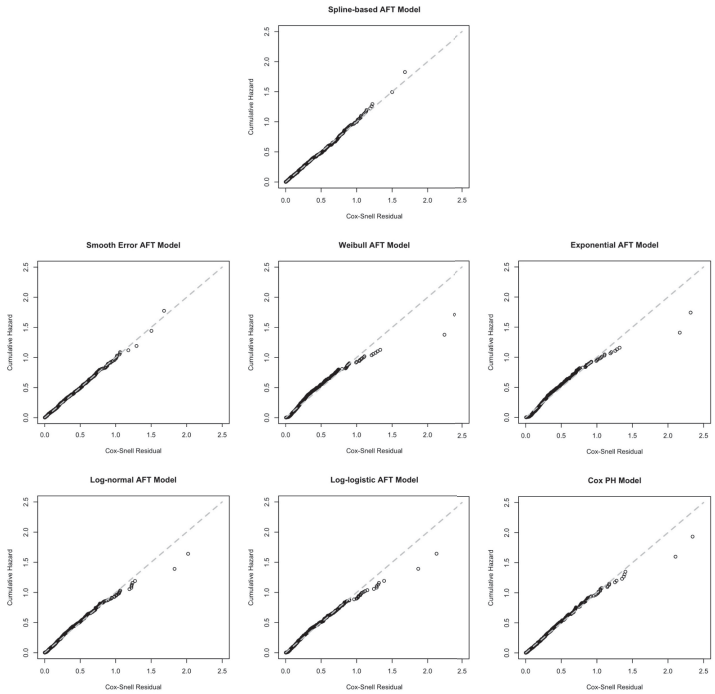


Figure A-8: Cox-Snell residual plots for alternative models in the colon cancer study.

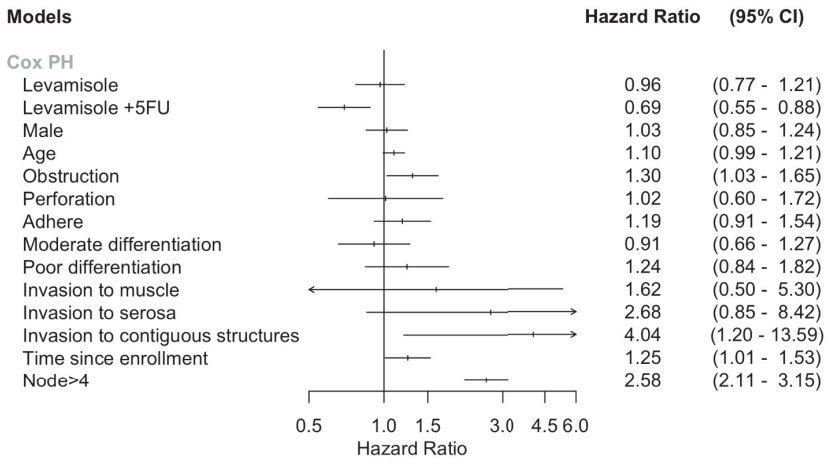


Figure A-9: Estimated covariate effects from the Cox PH model in the colon cancer study.

The survival curves estimated from the proposed spline-based AFT model are shown in Figure A-10 for the control, Levamisole and Levamisole+5FU groups, assuming mean age and all other covariates are at the reference level.

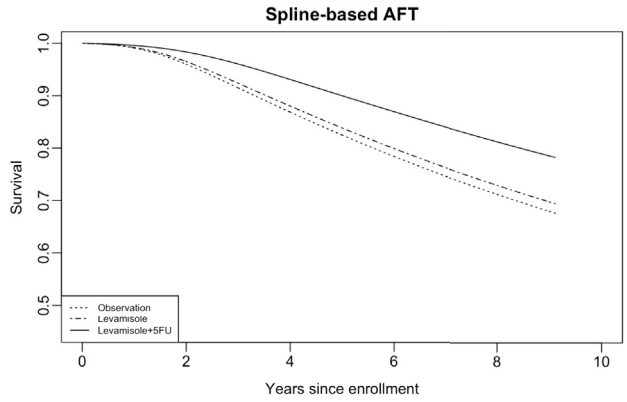


Figure A-10: Survival curves estimated from the spline-based AFT model comparing patients at mean age treated with Levamisole alone, Levamisole+5FU and untreated control group. All other covariates are assumed to be the reference level.

In Figure A-11, we show the survival curve, using the results from the spline-based AFT model, for four random selected patients in this study. Characteristics of these four patients are provided in Table A-9. This allows us to estimate/predict the survival curve for new patients with their covariate information measured at the baseline. Time-specific survival probability, as shown in Table A-10, can be calculated for each treatment option for patients who have the same covariate pattern with Patient A-D. This, as well as the survival curves, can potentially help inform treatment decision. For example, given limited health care source, clinician may decide which patient should receive the treatment with Levamisole+5FU with priority.

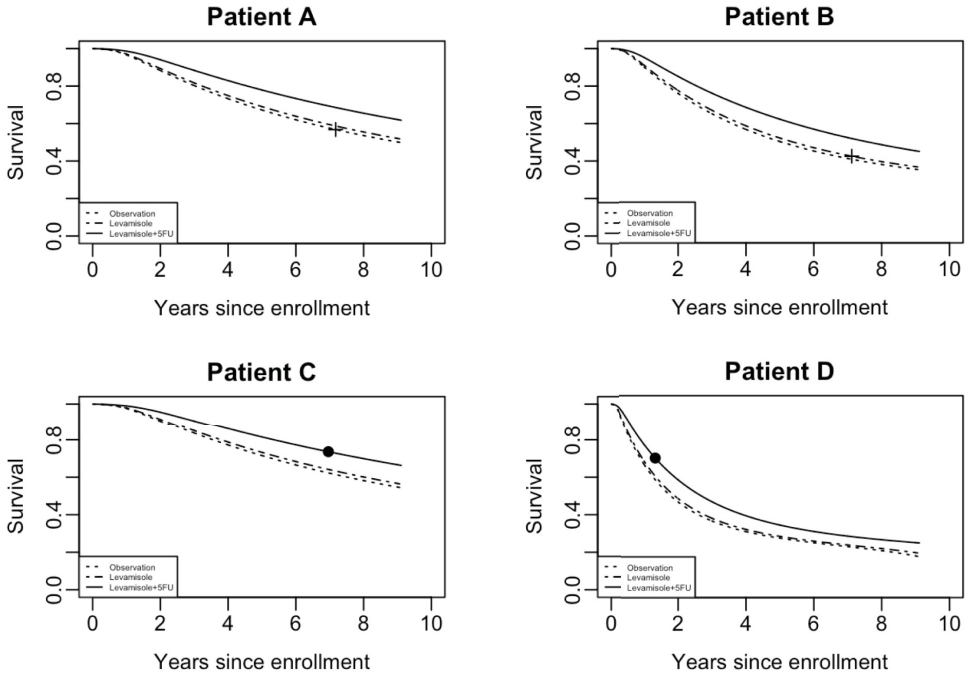


Figure A-11: Survival curves estimated from the spline-based AFT model for four random selected patients in the stage III colon cancer study. (Characteristics of the four patients are listed in Table A-9). The cross symbol (+) represents the observed censoring time and the dot symbol (•) represents the observed time of death. The survival curve in which the symbol is located represents the estimated survival for the treatment actually received, while the other curve corresponds to the survival had the patient received the alternative treatments.

Table A-9: Characteristics of the four random selected patients in the stage III colon cancer study

	Patient A	Patient B	Patient C	Patient D
Age	58	61	70	42
Sex	Female	Male	Male	Female
Treatment	Observation	Levamisole	Levamisole+5FU	Levamisole+5FU
Obstruction	No	No	No	No
Perforation	No	No	No	No
Adherence to nearby organs	No	No	No	No
Histologic differentiation	Moderately well	Moderately well	Moderately well	Poor
Depth of invasion	Serosa	Contiguous structures	Muscle	Serosa
Days from surgery to cohort entry	7-20	7-20	21-35	21-35
Positive lymph nodes >4	No	No	No	Yes

Table A-10: Five-year survival probability under different treatment for new patients who have the same covariate pattern with Patient A-D

Covariate Pattern	Observation only	Levamisole alone	Levamisole+5FU
Patient A	0.67	0.69	0.78
Patient B	0.50	0.52	0.62
Patient C	0.72	0.73	0.81
Patient D	0.28	0.29	0.35

Appendix B: Appendix to Manuscript 2

B.1 Assumptions under the AFT model

B.1.1 Demonstration of the constant time ratio assumption

For simplicity, Figure B-1 is generated with a constant baseline hazard function ($\lambda_0 = 1$) for the subgroup with $X = 0$. X is a protective factor that decelerates the event course by a constant time ratio, i.e., $e^{\beta} = \frac{2}{3}$ ($T(X = 1) = 1.5T(X = 0)$). The event times follow the exponential distributions with the corresponding rates. Therefore, the hazard function and the survival function when $X = 0$ are given by, respectively, $\lambda(t|X = 0) = \lambda_0$ and $S(t|X = 0) = \exp(-\lambda_0 t)$.

Under the AFT model, the hazard function and the survival function when $X = 1$ are given by, respectively, $\lambda(t|X = 0) = \lambda_0 e^{\beta}$ and $S(t|X = 1) = \exp(-\lambda_0 e^{\beta} t)$. Now we calculate the event times for any given value of a survival probability s in both groups. The times for attaining this probability when $X = 0$ and $X = 1$ are expressed as, $T(s|X = 0) = -\frac{1}{\lambda_0}\log(s)$ and $T(s|X = 1) = -\frac{1}{\lambda_0 e^{\beta}}\log(s)$. Thus, we have $\frac{T(s|T=1)}{T(s|X=0)} = e^{-\beta} = 1.5$. It clearly demonstrates that the time ratio is a constant during the entire follow-up, such that any quantiles of survival time are prolonged by 1.5 times in the subgroup with $X = 1$.

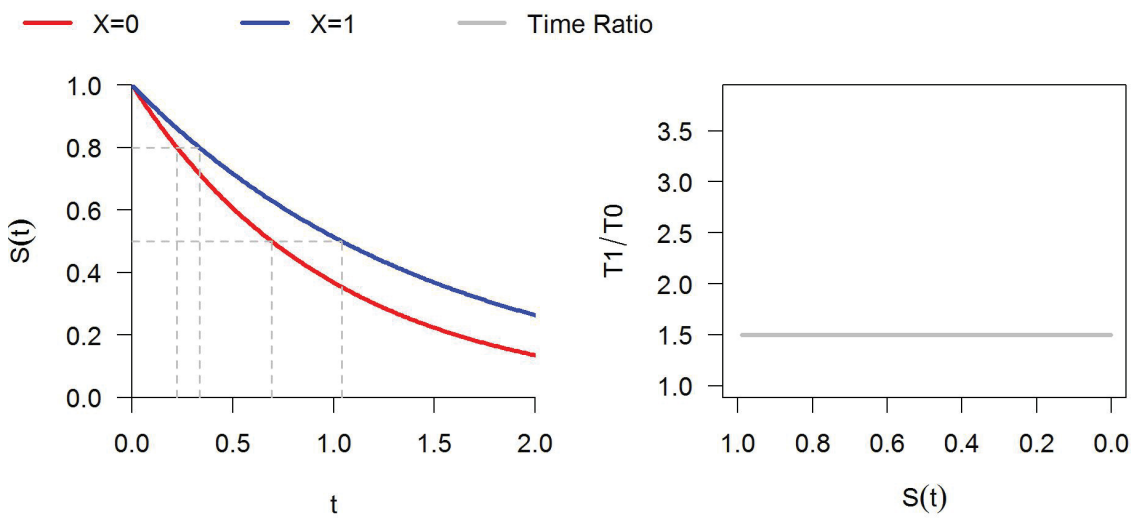


Figure B-1: Survival curves (left panel) and time ratio (right panel) under the constant time ratio assumption for a protective binary covariate X under an exponential AFT model ($e^{\beta} = \frac{2}{3}$). The dashed lines in the left panel show the corresponding median and 80% percentile of the survival times in both subgroups.

B.1.2 Demonstration of the violation of the constant time ratio assumption

Figure B-2 is again generated with a constant baseline hazard function (λ_0) for the subgroup with $X = 0$, so the event time for this subgroup is again assumed to follow the exponential distribution. In this setting, X is a protective factor, however, to relax the constant time ratio assumption, we assume that $X = 1$ decelerates the event time by a time-dependent time ratio, i.e., $e^{\beta t} = \frac{1}{1.5^t}$. The hazard function and the survival function when $X = 0$ remain the same in the previous setting B.1.1 with constant time ratio. However, under this AFT model, the event time in the subgroup with $X = 1$ does not follow exponential distribution. The hazard function and the survival function when $X = 1$ are respectively derived as,

$$\lambda(t|X = 1) = \lambda_0 e^{\beta t},$$

and

$$S(t|X=1) = \exp\Big(-\int_0^t \lambda_0 e^{\beta u} du\Big) = \exp\Big(-\frac{\lambda_0}{\beta}(e^{\beta t}-1)\Big).$$

It follows that for any given value of a survival probability s, the times for attaining this probability when $X = 0$ and $X = 1$ are given by, $T(s|X=0) = -\frac{1}{\lambda_0}\log(s)$ and $T(s|X=1) = \frac{1}{\beta}\log\Big(1-\frac{\beta}{\lambda_0}\log(s)\Big)$.

Apparently, the time ratio $\frac{T(s|X=1)}{T(s|X=0)}$ is not a constant and depends on the survival probability being evaluated. When there are only 20% subjects who had event in both groups, the time ratio is 1.05 (treated vs. control), whereas when there are 50% subjects who had event in both groups, the time ratio is 1.17.

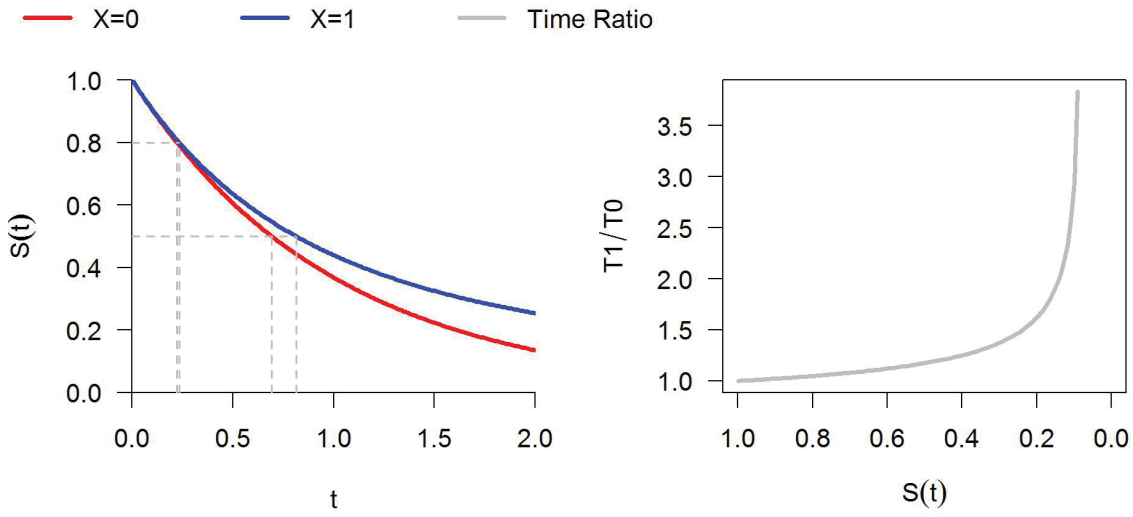


Figure B-2: Survival curves (left panel) and time ratio (right panel) for a protective binary covariate X under an exponential hazard function with time-dependent time ratio. The dashed lines in the left panel show the corresponding median and 80% percentile of the survival times in both subgroups.

B.1.3 Violation of the linearity assumption under AFT models

Violation of linearity assumption implies that the time ratio for one unit increase in X depends on the value of X. For example, the ratio of survival times, associated with a 10-year increase in age, may differ when comparing two subgroups who are aged (a) 70-year old vs. 60-year-old, and (b) 30-year old vs. 20-year old.

B.2 Details of methods

B.2.1 Alternating conditional estimation (ACE) algorithm

We adapt an iterative alternating conditional estimation (ACE) algorithm [33, 97] to estimate the parameters of our flexible extension of the AFT model (5.3.5). In particular, our ACE algorithm divides the parameter space into three mutually exclusive subsets, consisting of the spline coefficients for, respectively: (i) the baseline hazard $\gamma_k, k = 1, \cdots, K$, (ii) the TD effects: $b_{q,j}, q = 1, \cdots, Q$, for all covariates, and (iii) the NL effects: $a_{l,j}, l = 1, \cdots, L$, for continuous covariates. The algorithm then iterates across the following three consecutive steps, each involving estimating only one of the above subsets:

(1) Estimate $\boldsymbol{\gamma}$ for the log hazard function $\lambda_0\Big(\exp\Big(\sum_j \beta_j(t) g_j(X_j)\Big)t\Big)$ in equation (5.3.8) of the main manuscript, conditional on the estimates, obtained from the previous iteration, of both time-dependent effect $\beta_j(t)$ in equation (5.3.6) and non-linear function $g_j(X_j)$ in equation (5.3.7);

(2) Estimate $\boldsymbol{b}$ for the time-dependent effect $\beta_j(t)$, conditional on the estimates of $\boldsymbol{\gamma}$ from step (1) of the same iteration and $g_j(X_j)$ from step (3) of the previous iteration;

(3) Estimate $\boldsymbol{a}$ for the non-linear function $g_j(X_j)$, conditional on the estimates of $\boldsymbol{\gamma}$ and $\boldsymbol{b}$ from steps (1) and (2) of the same iteration.

In the step (1) of the first iteration, we fit a 'naïve' multivariable exponential AFT model, assuming a constant baseline hazard $\hat{\lambda}^{(0)}$, constant time ratio ($\hat{\beta}_j^{(0)}$) over time for variable X_j and linear effects for all continuous variables. We then set initial values in equation (5.3.8) of the main manuscript to the estimated hazard rate: $\gamma_k^{(0)} = \hat{\lambda}^{(0)}$, $k = 1, \cdots, K$, and set initial values in equation (5.3.7) for each covariate X_j to the estimated constant time ratio: $b_{q,j}^{(0)} = \hat{\beta}_j^{(0)}$, $q = 1, \cdots, Q$, $j = 1, \cdots, J$. For a continuous covariate X_j, we fit a linear regression using the X_j as the dependent variable and the spline basis in equation (5.3.6), $A_{l,j}(X_j)$, as independent variables. The coefficients obtained from this linear regression are used as initial values for $a_{l,j}^{(0)}$. The normalization property of B-splines ensures that $\sum_{k=1}^{K} S_k(\cdot) = \sum_{l=1}^{L} A_{l,j}(X_j) = \sum_{q=1}^{Q} B_q(t) = 1$, thus these initial values imply $\lambda_0\big(\exp\Big(\sum_j \beta_j(t) g_j(X_j)\Big)t\big) = \hat{\lambda}^{(0)}$, $\beta_j(t) = \hat{\beta}_j^{(0)}$ and $g_j(X_j) = \hat{X}_j$. In the step (2) of the first iteration, the initial value of $b_{q,j}$ is again set to be $b_{q,j} = \hat{\beta}_j^{(0)}$, $q = 1, \cdots, Q$, $j = 1, \cdots, J$, while conditional on the initial value $a_{l,j}^{(0)}$ and $\hat{\boldsymbol{\gamma}}$ estimated from step (1). Likewise, in the step (3) of the first iteration, the initial value of $a_{l,j}$ is set to be $a_{l,j}^{(0)}$ while conditional on $\hat{b}_{q,j}$ estimated from step (2) and $\hat{\boldsymbol{\gamma}}$ estimated from step (1). In the subsequent iterations of the algorithm, the initial values for the parameters being estimated at each step are set to be the corresponding estimates obtained at the previous iteration.

The estimation process is repeated iteratively until the difference between the log-likelihood from two consecutive iterations is less than a pre-specified threshold, set to be 10^{-5} in our simulation and analyses reported in this article. R programs for implementing the complex log-likelihood function in equation (5.3.9) of the manuscript and the ACE algorithm are provided to achieve the estimation of the hazard function, NL and TD effects. For the estimation in each step, our program employs the BFGS quasi-Newton method implemented by 'optim' function in R, using the first derivatives of the log-likelihood function in equation (5.3.9), with respect to each of the three subsets of the parameters. Appendix B.2.3 describes why and how, the final TD and NL estimates for the same continuous covariate are re-scaled to facilitate interpretation of result.

The first derivative of the log-likelihood with respect to γ

For the r^{th} spline coefficient in the estimation of the hazard, the first derivative of the log-likelihood is given by:

$$\frac{\partial \log L}{\partial \gamma_r} = \sum_{i=1}^{n} \left[\delta_i S_r(w_i) - \int_0^{t_i} \exp\left(\sum_{j=1}^{J} \beta_j(u) g_j(X_{ij})\right) \exp\left(\sum_{k=1}^{K} \gamma_k S_k(w_i')\right) S_r(w_i') du \right]$$

where $w_i = \exp\left(\sum_{j=1}^{J} \beta_j(t_i) g_j(X_{ij})\right) t_i$, $w_i' = \exp\left(\sum_{j=1}^{J} \beta_j(u) g_j(X_{ij})\right) u$, $\beta_j(u) = \sum_{q=1}^{Q} b_{q,j} B_q(u)$, and $g_j(X_{ij}) = \sum_{l=1}^{L} a_{l,j} A_{l,j}(X_{ij})$; $S_r(\cdot)$, $A_{l,j}(X_{ij})$ and $B_q(u)$ are B-spline basis for the modeling of the hazard, NL and TD effects, respectively.

The first derivative of the log-likelihood with respect to *b*

For the r^{th} spline coefficient for the estimation of the TD effect of covariate j, the first derivative of the log-likelihood is given by:

$$\frac{\partial \log L}{\partial b_{r,j}} = \sum_{i=1}^{n} \left[\delta_i B_r(t_i) g_j(X_{ij}) - \int_0^{t_i} \exp\left(\sum_{j=1}^{J} \beta_j(u) g_j(X_{ij})\right) \exp\left(\sum_{k=1}^{K} \gamma_k S_k(w_i')\right) B_r(u) g_j(X_{ij}) du \right]$$

The first derivative of the log-likelihood with respect to *a*

For the r^{th} spline coefficient for the estimation of the NL effect of covariate j, the first derivative of the log-likelihood is given by:

$$\frac{\partial \log L}{\partial a_{r,j}} = \sum_{i=1}^{n} \left[\delta_i \beta_j(t_i) A_{r,j}(X_{ij}) - \int_0^{t_i} \exp\left(\sum_{j=1}^{J} \beta_j(u) g_j(X_{ij})\right) \exp\left(\sum_{k=1}^{K} \gamma_k S_k(w_i')\right) \beta_j(u) A_{r,j}(X_{ij}) du \right]$$

B.2.2 Bootstrap-based confidence intervals

Standard large-sample inference may not accurately quantify the sampling variance of the estimates, which resulted from the three-step iterative conditional estimation of parameters in the complex modeling that consist of (i) hazard functions, (ii) TD effects and (iii) NL effects

[41]. Therefore, we rely on non-parametric bootstrap procedure [118] to assess the stability of the flexible AFT model. In each of M bootstrap resamples, we use the proposed ACE algorithm to estimate the proposed flexible AFT model, adjusting for potential TD and NL effects. The pointwise 95% confidence bands around the estimated TD and NL functions for covariate j, $\beta_j(t)$ and $g_j(X_j)$, are obtained by connecting the 2.5^{th} and 97.5^{th} percentiles of the empirical distribution of the corresponding estimates across the M resamples, for each t and for each specific value of the covariate X_j, respectively. The confidence interval bands for the hazard function, survival curve conditional on individual covariate pattern as well as the time-dependent time ratio can be obtained likewise.

B.2.3 Rescaling of the NL and TD effects

The ACE algorithm ensures that the TD and NL functions, $\beta_j(t)$ and $g_j(X_j)$, that are multiplied by each other in equation (5.3.5) of the manuscript, are estimated separately, respectively, in steps (2) and (3) of each iteration of ACE. However, the two functions may share an arbitrary scale factor, i.e., for any $\xi_j \neq 0$, $\beta_j(t)g_j(X_j)$ can be represented as $\xi_j\beta_j(t)\frac{g_j(X_j)}{\xi_j}$. In the final estimates, where we may obtain $\tilde{\beta}_j(t)$ as the estimate of $\xi_j\beta_j(t)$ and $\tilde{g}_j(X_j)$ as the estimate of $\frac{g_j(X_j)}{\xi_j}$, it is infeasible to disentangle such scale factor ξ_j. To describe the TD and NL effects in a meaningful manner and evaluate the estimation of $\beta_j(t)$ and $g_j(X_j)$ in simulation studies using their corresponding true functions as benchmarks, we propose to present the scaled version of the TD and NL effects using the range of the NL effect. Specifically, we present the estimated TD effect using $\hat{\beta}_j(t) = \tilde{\beta}_j(t) \times \left[\max(\tilde{g}_j(X_j)) - \min(\tilde{g}_j(X_j))\right]$. $\hat{\beta}_j(t)$ can be interpreted as the adjusted log TD effect, at time t, between the values of X_j corresponding to, respectively, the highest and the lowest risks. Accordingly, we present the NL effect using $\hat{g}_j(X_j) = \tilde{g}_j(X_j)/\left[\max(\tilde{g}_j(X_j)) - \min(\tilde{g}_j(X_j))\right]$. $\hat{\beta}_j(t)$, so that the NL effect is constrained to be between 0 and 1. ξ_j may be negative, in which case, we rely on some prior knowledge about the NL effect of X_j to obtain the correct direction of both effects. For example, we may

know *a priori* that the risk of event associated with X_j, such as age and SOFA in our septic shock study, at 90% percentile is larger than that at 10% percentile. If this is true from the results obtained with ACE, i.e., $\tilde{g}_j(X_{j(0.9)}) > \tilde{g}_j(X_{j(0.1)})$, implying $\xi_j > 0$, then we present $\hat{\beta}_j(t)$ and $\hat{g}_j(X_j)$, otherwise, we present $-\hat{\beta}_j(t)$ and $-\hat{g}_j(X_j)$. Note that the overall effect of X_j at a given time t on the log hazard is the product of $\beta_j(t)$ and $g_j(X_j)$ jointly. Both $\hat{\beta}_j(t)\hat{g}_j(X_j)$ and $\tilde{\beta}_j(t)\tilde{g}_j(X_j)$ give the same estimation of $\beta_j(t)g_j(X_j)$.

B.3 Details of the simulation studies

B.3.1 Data generation procedure

For each scenario, we simulated 100 independent random samples. For each random sample, data were generated by the following steps:

(1) Generate N covariate vectors. In both scenarios, X_1 was generated from a Bernoulli distribution with probability 0.5. In scenario 1, both X_2 and X_3 were generated from a standard normal distribution ($\mathcal{N}(0, 1)$), whereas in scenario 2, they were generated from a uniform distribution ($U[-1, 1]$). The three covariates were independently generated in both scenarios.

(2) Generate hypothetical event times, conditional on covariate vectors generated in step (1). For each covariate vector, we first calculated the values of the true survival function $S(t|X_1, X_2, X_3)$ consistent with the true model expressed in equation (5.4.1). Due to the complexity of our model, no closed form could be derived for $S(t|X_1, X_2, X_3)$. Therefore, we adopted a numerical grid search approach similar to that employed in simulations evaluating full maximum likelihood estimation of a flexible extension of the PH model [41]. Specifically, we first calculated the consecutive survival probabilities corresponding to discrete times t, increased in increments of 3×10^{-4} year, from 0 to 6 years of follow-up, using numerical integration $S(t|X_1, X_2, X_3) = \exp\left(-\int_0^t \lambda(u|X_1, X_2, X_3)du\right)$.

We then generated a random value u from a uniform distribution $U[0, 1]$, and searched for the time t such that the difference between the corresponding true $S(t|X_1, X_2, X_3)$ and u was less than a threshold set to be 10^{-3}. The resulting t was considered to represent the expected event time for a subject with the corresponding covariate vector.

(3) Generate censoring times and the observed times. Administrative right censoring at $C_1 = 6$ years was imposed by the simulation design. Additionally, uniformly distributed random censoring times (C_2) were generated to mimic losses to follow-up, so as to achieve about 80% overall event rate (20% censoring) in both scenarios. An individual's observed time was then determined as $\min(t, C_1, C_2)$, and the subject's status was defined as ($\delta_i = 1$) if it corresponded to the event time t generated in step (2).

B.3.2 Results of the estimated TD and NL effects in alternative simulation scenarios

The gray curves are the individual estimates from 100 samples, and the pointwise mean is shown by the white curve. The black dashed curve represents the true rescaled NL and TD functions.

Alternative simulation 1: sample size: $N = 1,000$, event rate: 40%

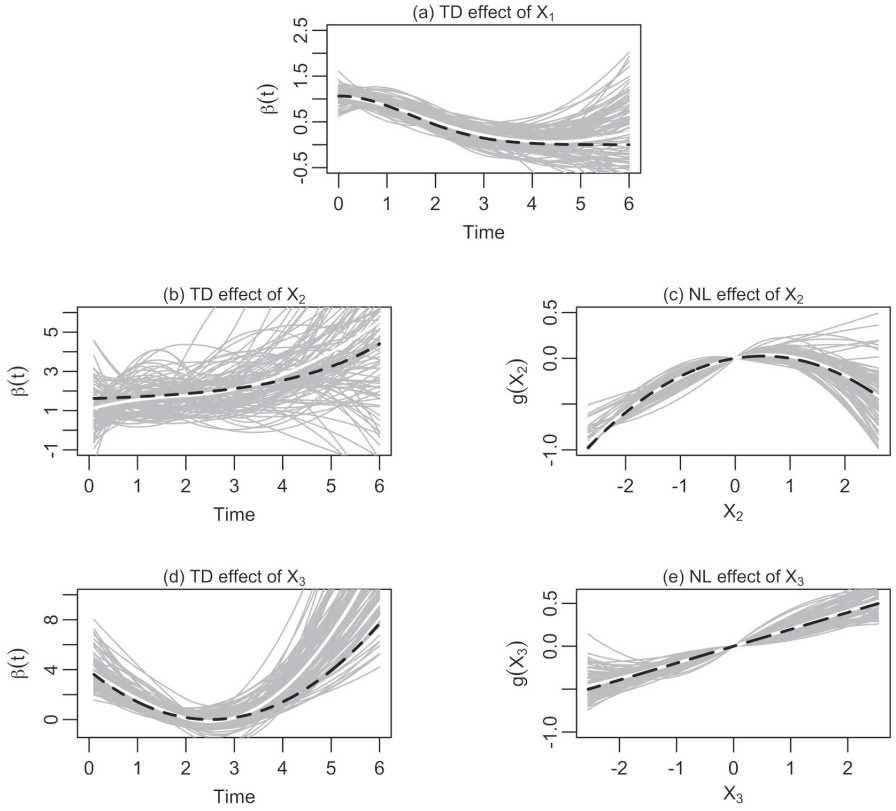


Figure B-3: Results of the estimated TD and NL effects using 100 samples in alternative simulation 1 (sample size: $N = 1,000$, event rate: 40%).

Alternative simulation 2: sample size: N = 650, event rate: 80%

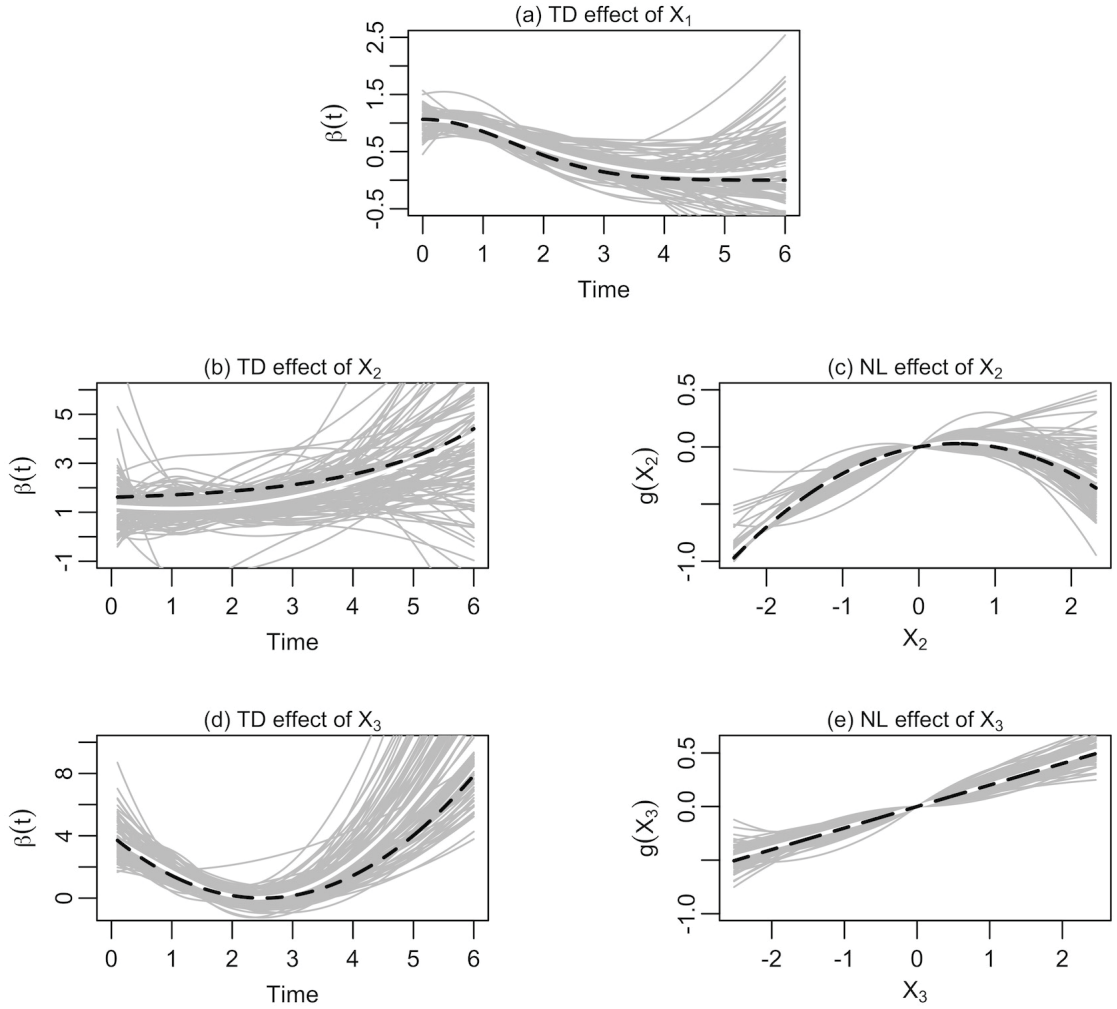


Figure B-4: Results of the estimated TD and NL effects using 100 samples in alternative simulation 2 (sample size: N = 650, event rate: 80%).

Alternative simulation 3: sample size: N = 650, event rate: 40%

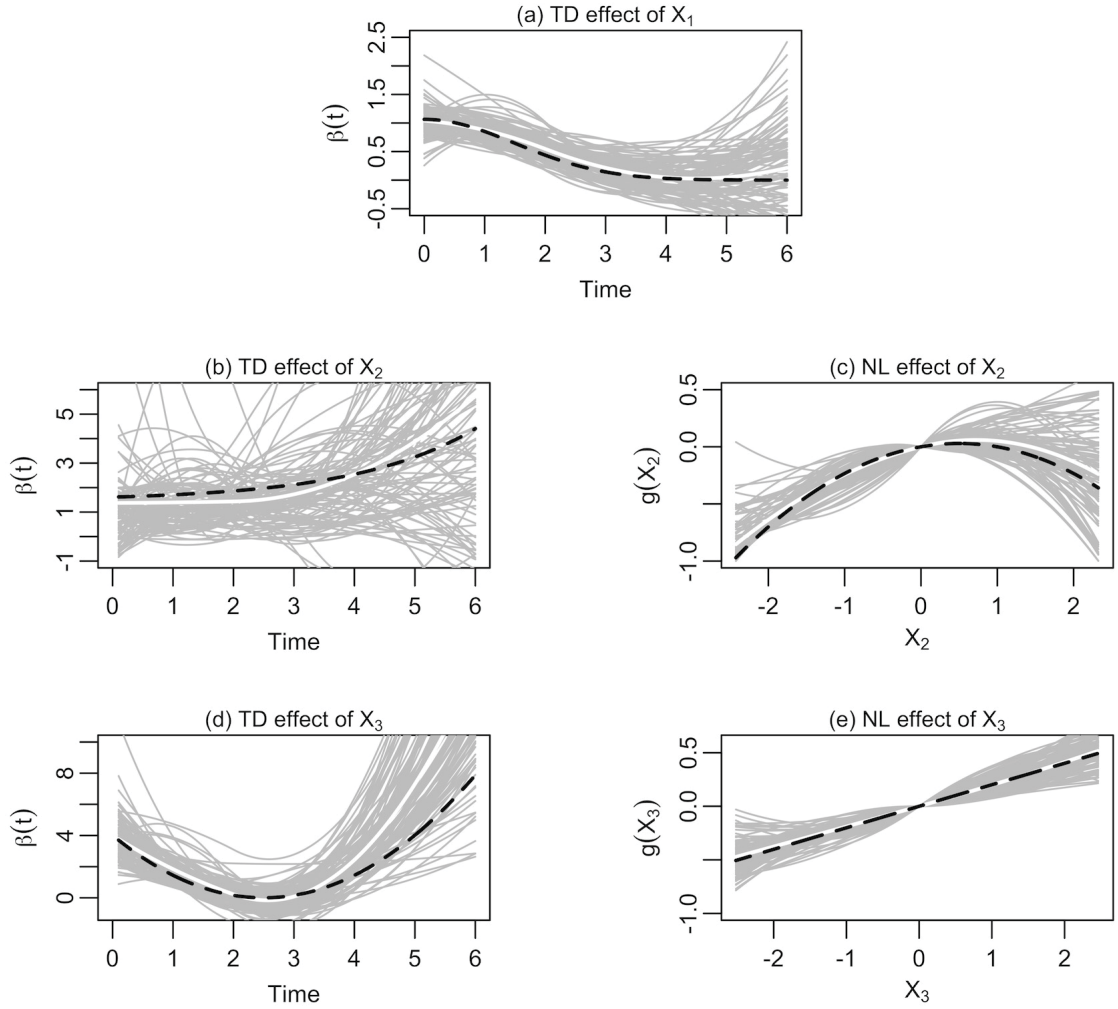


Figure B-5: Results of the estimated TD and NL effects using 100 samples in alternative simulation 3 (sample size: N = 650, event rate: 40%).

Alternative simulation 4: sample size: $N = 1,500$, event rate: 80%

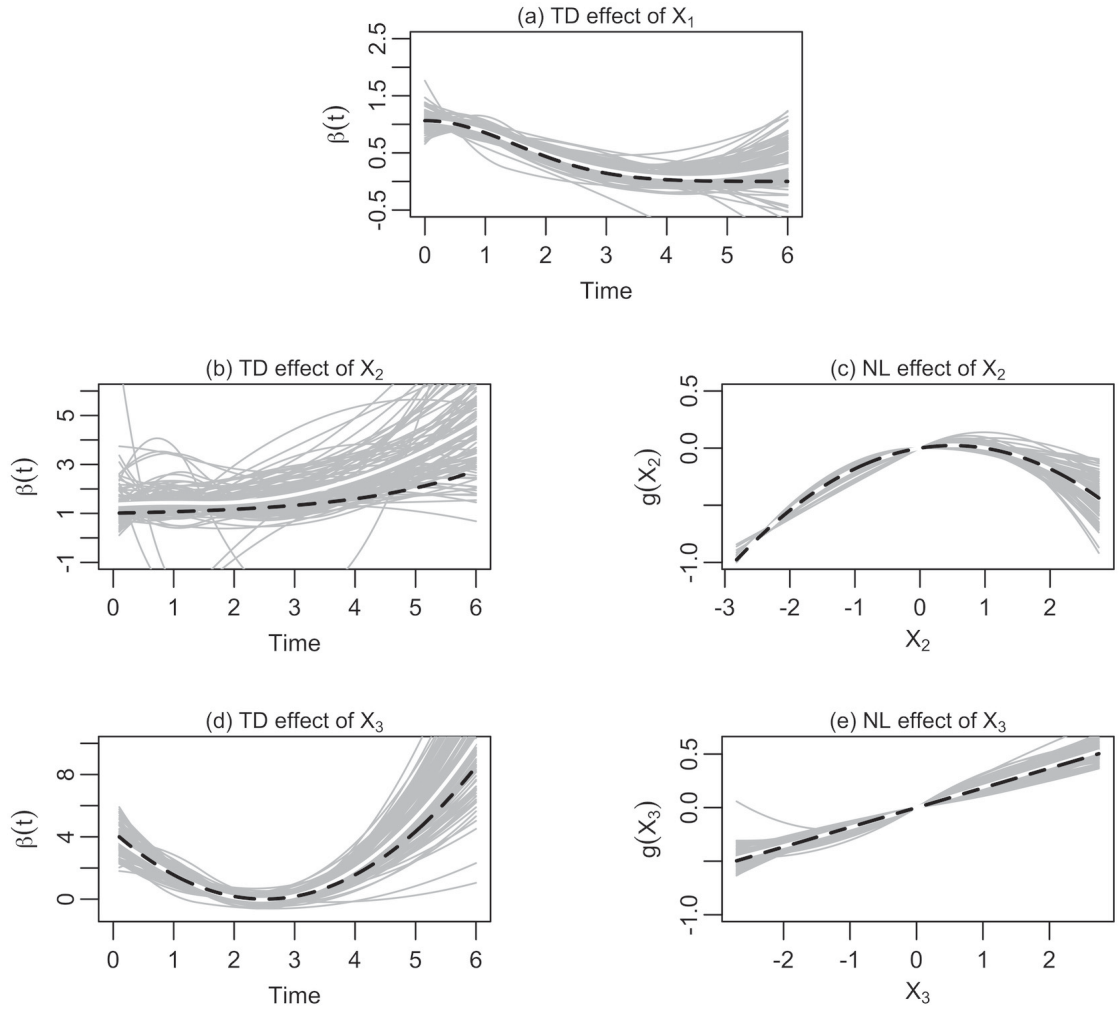


Figure B-6: Results of the estimated TD and NL effects using 100 samples in alternative simulation 4 (sample size: $N = 1,500$, event rate: 80%).

Alternative simulation 5: sample size: $N = 1,500$, event rate: 40%

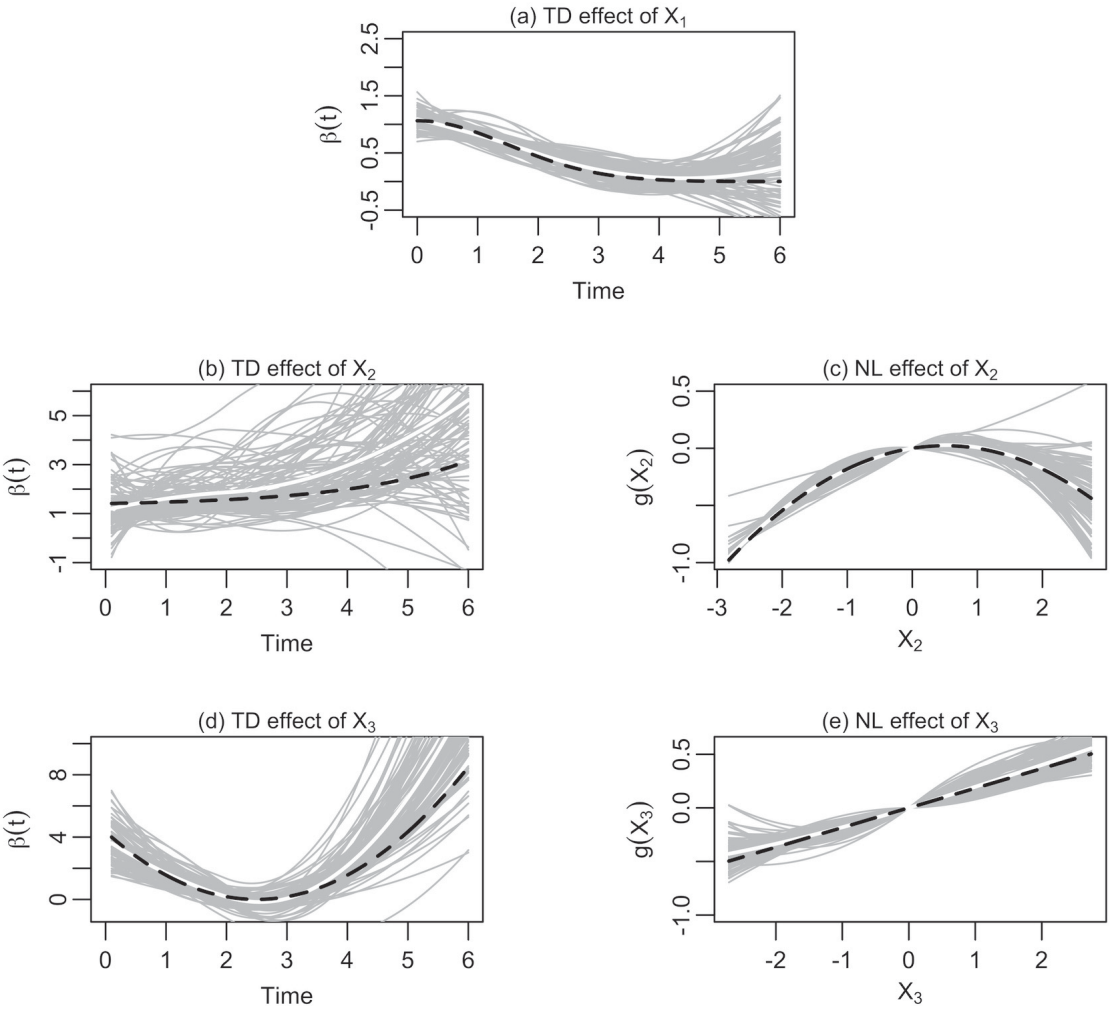


Figure B-7: Results of the estimated TD and NL effects using 100 samples in alternative simulation 5 (sample size: $N = 1,500$, event rate: 40%).

B.3.3 Results from model (i) and (ii) in the main simulation scenarios 1 and 2

Figure B-8 attempts to provide some insights regarding the implications of model mis-specification, with respect to covariate effects. In particular, it permits assessing the impact of ignoring both TD and NL effects by the 'conventional' Weibull AFT model (i), and ignoring the TD effects by the 'extended' Weibull AFT model (ii).

Figure B-8 shows the estimated constant time ratios from the 'conventional' Weibull AFT model (i) for scenarios 1 and 2, and the constant-over-time effects of X_1 and the NL effects of X_2 and X_3 estimated from the 'extended' Weibull AFT model (ii). Notice that in model (ii) that incorporates non-linear covariate effects, the effect of X_2 and X_3 are represented by $g_2(X_2)$ and $g_3(X_3)$ only, which are *a priori* constrained to be constant over time. In this case, conventional regression coefficients are used as the spline coefficients to construct the estimated NL functions. However, the TD effects can be characterized by neither (i) the 'conventional' Weibull AFT nor (ii) the 'extended' Weibull AFT models. Biased constant time ratios for all the three covariates are resulted from model (i) in both scenarios, while none of the NL effects are being estimated (the first and third row in Figure B-8). On the other hand, most of the estimated NL effects by model (ii) are biased (second and fourth row 4), although surprisingly, the NL functions are accurately recovered for X_2 in scenario 1. Overall, both models (i) and (ii) provide unsatisfactory results in terms of the estimation of the covariate effects.

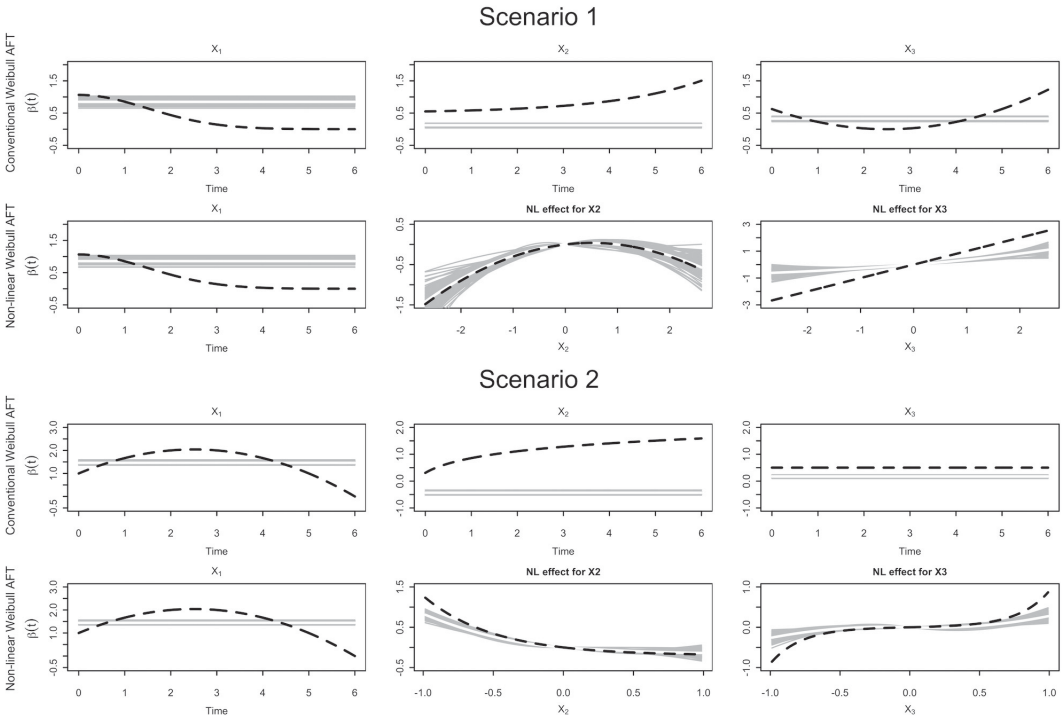


Figure B-8: Results of the estimated constant effects for all three covariates using 100 samples in simulation scenarios 1 and 2, provided by the 'conventional' Weibull AFT model (i) (first and third row) and the constant effect for X_1 and NL effects for X_2 and X_3 provided by the 'extended' Weibull AFT model (ii) (second and fourth row).The gray curves are the individual estimates from 100 samples, and the pointwise mean is shown by the white curve. The black dashed curve represents the true NL and TD functions.

B.3.4 Results of the estimated hazard function and survival curve in the main simulation scenarios 1 and 2

The hazard function and survival curve conditional on any given covariate pattern can be derived from all the three alternative models for both main scenarios. In addition to the setting ($X_1 = X_2 = X_3 = 0$) presented in the main manuscript, the results of two more covariate patterns are shown in this section.

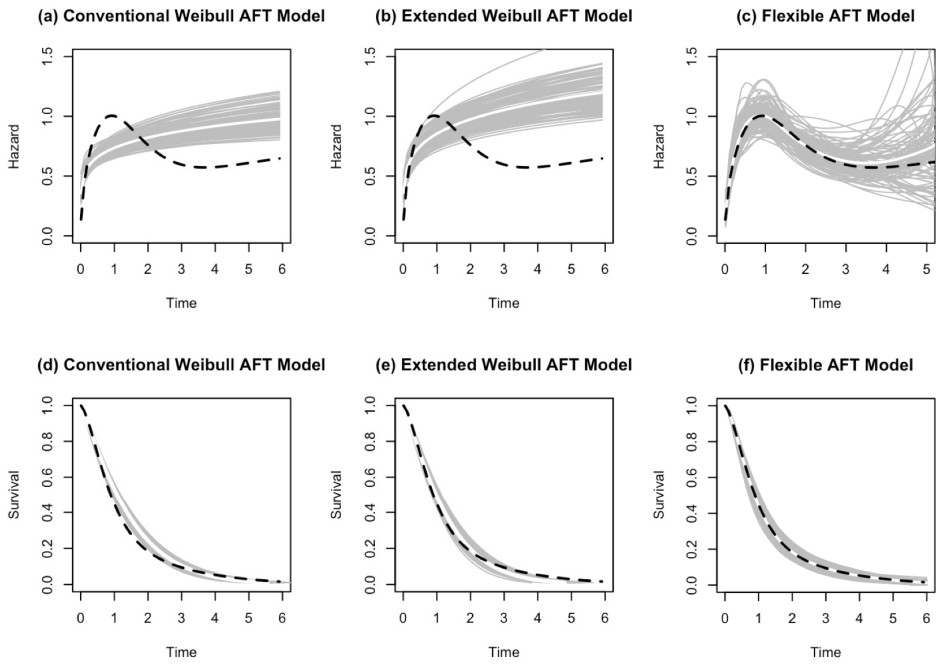


Figure B-9: The comparison of the estimation of hazard function and survival curve in the main simulation scenario 1, conditioning on the covariate pattern $X_1 = 1$, $X_2 = X_3 = 0$.

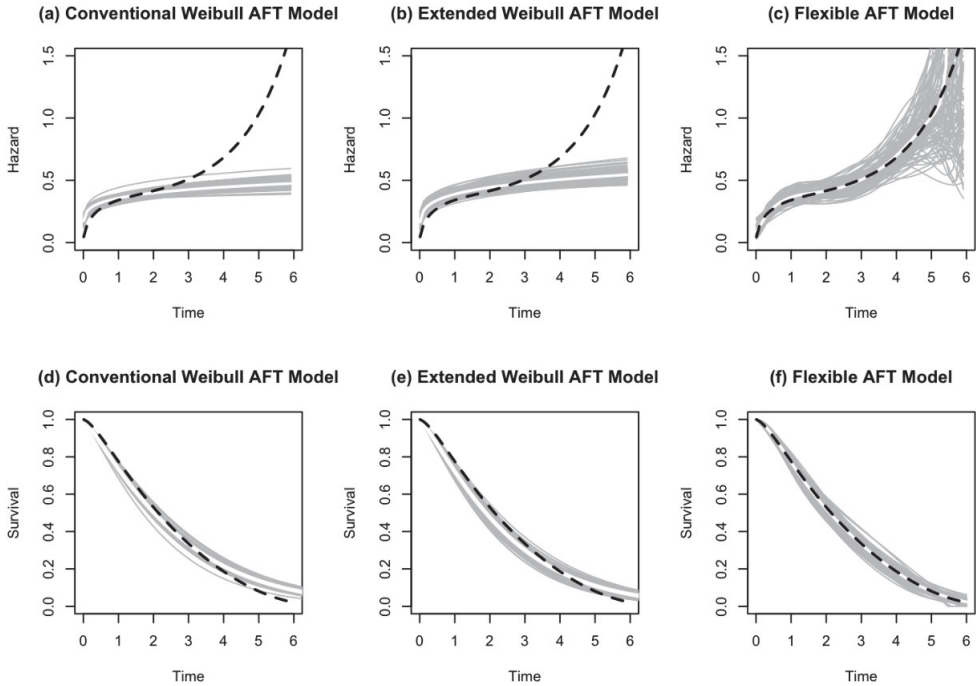


Figure B-10: The comparison of the estimation of hazard function and survival curve in the main simulation scenario 1, conditioning on the covariate pattern $X_1 = 1$, $X_2 = X_3 = 0.5$.

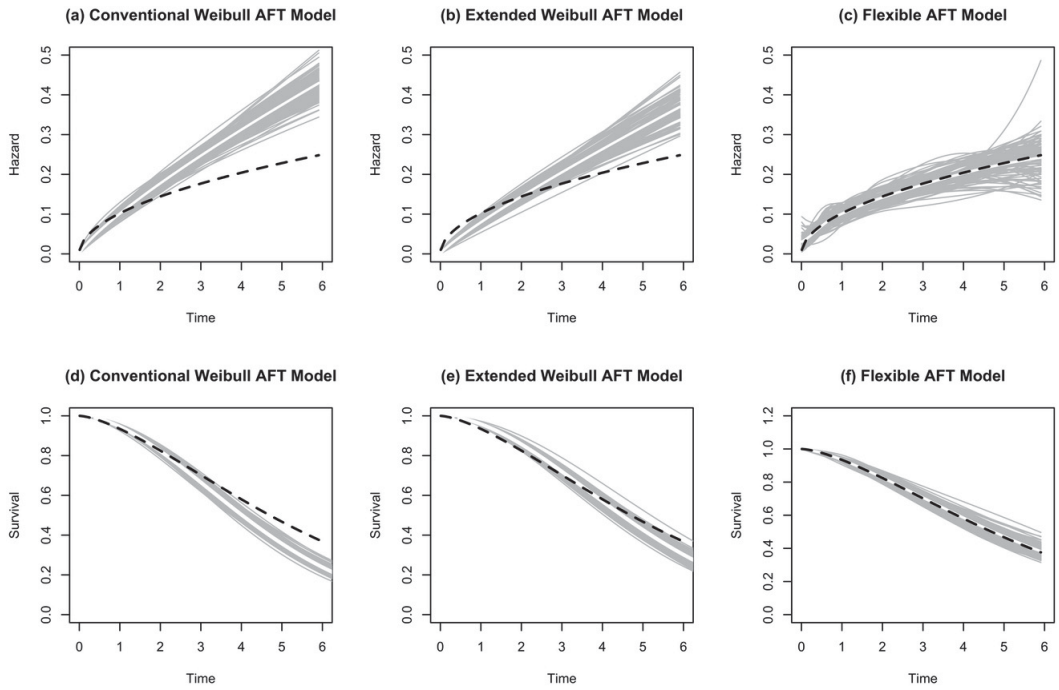


Figure B-11: The comparison of the estimation of hazard function and survival curve in the main simulation scenario 2, conditioning on the covariate pattern $X_1 = X_2 = X_3 = 0$.

B.3.5 Additional results of the estimated log time ratios in the main simulation scenarios 1 and 2

Two covariate patterns are shown in the labels on the top of each panel, along with the true survival times in both groups corresponding to specific q-quantile of the survival time.

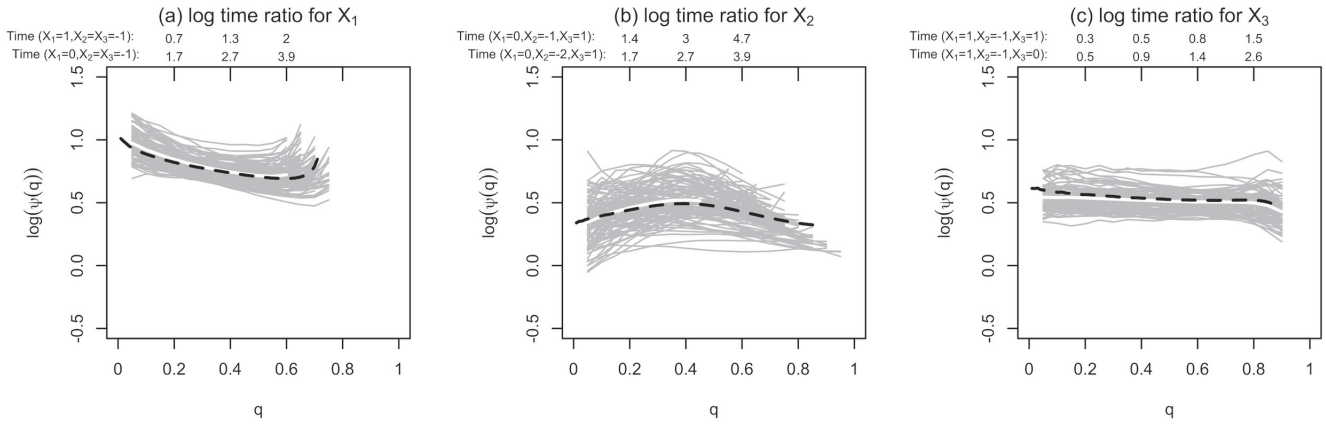


Figure B-12: Results of the estimated log time ratios by the flexible AFT model using 100 samples in simulation scenario 1, comparing two covariate patterns for each covariate.

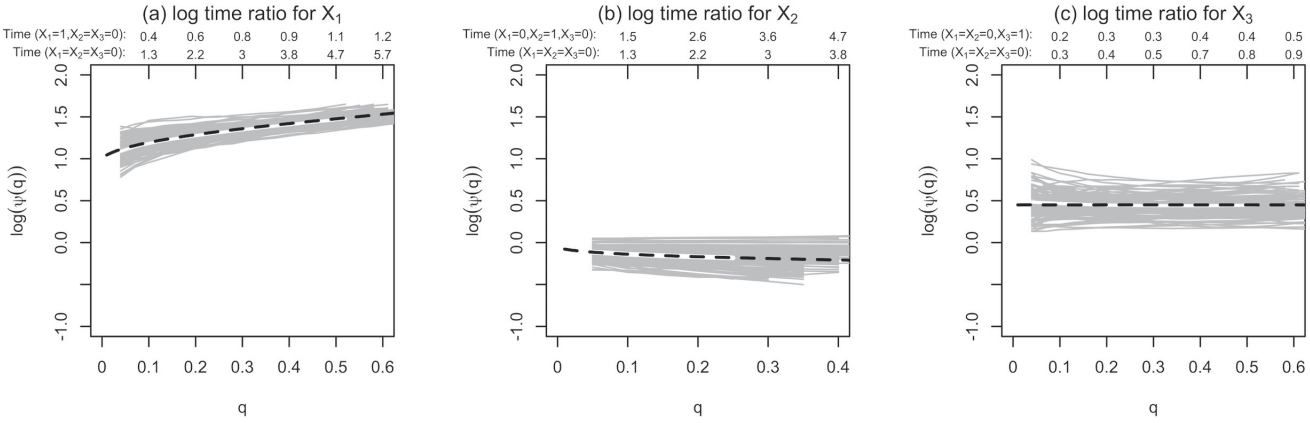


Figure B-13: Results of the estimated log time ratios by the flexible AFT model using 100 samples in simulation scenario 2, comparing two covariate patterns for each covariate.

B.3.6 Alternative simulation in univariate setting with constant time ratio and linear effect: $\beta(t)g(X) = 0.8 * X$

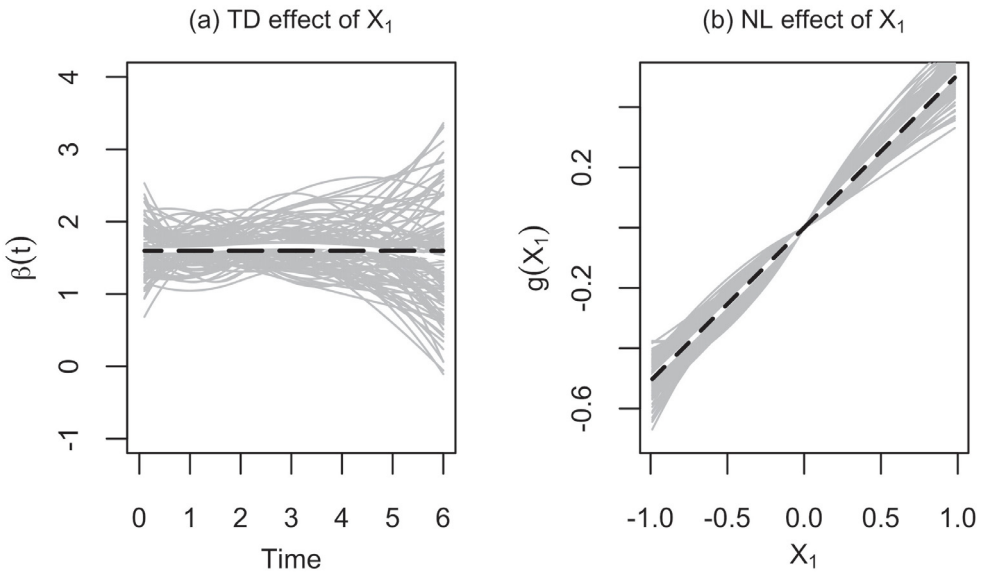


Figure B-14: Results of the estimated TD and NL effects in univariate setting with constant time ratio and linear effect.

Table B-1: Comparison of mean akaike information criterion (AIC) from three alternative models in simulation where constant time ratio and linear function are the true effects

	Flexible AFT model (df=14)	Conventional Weibull AFT model (df=3)	Non-linear Weibull AFT model (df=5)
Log-likelihood	-1355.071	-1359.705	-1358.772
AIC	2738.142	2725.411	2727.543

B.4 Details of the real-life application

B.4.1 Modeling strategy in the septic shock main analysis

In our main analyses we used a modeling strategy based on the proposed flexible AFT model, consisted of three stages. In stage (1), we considered the five aforementioned important

baseline prognostic factors, i.e, age, Sepsis-related Organ Failure Assessment (SOFA) score, immunosuppression, infection site and Knaus score. The initial flexible multivariable AFT model, estimated at stage (1), included the TD effects for all the five covariates, and the NL effects for both continuous variables: age and SOFA score. In stage (2), stepwise variable addition was performed to check if adding other prognostic factors (whether or not the germ was identified, infection type (community-acquired vs. nosocomial), cirrhosis status (yes vs. no) in the mode would improve the goodness-of-fit. In the first step, the TD effect of each of the three prognostic factors were added, one at a time to the initial model, the effect with the minimum AIC value was added. In the consecutive steps, we compared the models when adding the remaining TD effects. The stepwise addition continued until no further additions could improve the AIC. In stage (3), stepwise backward variable deletion was employed based on the model resulted from stage (2). In each step, the TD or the NL effect of each covariate was excluded, one at a time, and the effect whose deletion resulted in the largest reduction in AIC value was eliminated until no improvement in AIC was achieved. If a TD or NL effect was eliminated in one step, the covariate was still included in the model but conform with the constant time ratio or linearity assumption. The model resulted from stage (3) was the final flexible AFT model.

Table B-2: Results of each stage in building the final flexible AFT model in the septic shock study

	Covariates		Log-likelihood	Degree of freedom	AIC
Stage 1			-2065.971	34	4199.942
Selected covariates/ Initial Model	Immunosuppression (TD); Knaus score (TD); Infection site (TD); Age (TD+NL); SOFA(TD+NL)				
Stage 2	Stage 2				
Step 1		Addition			
	Germ	+TD	-2062.404	38	4200.808
	Cirrhosis	+TD	-2046.147	38	4168.294
	Infection type (nosocomial)	+TD	-2052.234	38	4180.468
Step 2		Addition			
	Germ	+TD	-2046.776	42	4177.551
	Infection type (nosocomial)	+TD	-2041.292	42	4166.584
Step 3		Addition			
	Germ	+TD	-2041.243	46	4174.487
Selected covariates	Immunosuppression (TD); Knaus score (TD); Infection site (TD); Cirrhosis (TD); Infection type (TD); Age (TD+NL); SOFA(TD+NL)				
Stage 3					
Step 1		Deletion			
	Immunosuppression	-TD	-2052.812	39	4183.623
	Knaus score	-TD	-2041.937	39	4161.874
	Infection site (urinary)	-TD	-2041.828	39	4161.657
	Cirrhosis	-TD	-2059.237	39	4196.474
	Infection type (nosocomial)	-TD	-2049.263	39	4176.527
	Age	-TD	-2153.189	38	4382.379
	SOFA score	-TD	-2060.938	38	4197.875
	Age	-NL	-2044.731	38	4165.461
	SOFA score	-NL	-2044.293	38	4164.586
Step 2		Deletion			
	Immunosuppression	-TD	-2045.444	36	4162.888
	Knaus score	-TD	-2043.001	36	4158.002
	Cirrhosis	-TD	-2060.006	36	4192.013
	Infection type (nosocomial)	-TD	-2045.215	36	4162.431
	Age	-TD	-2153.198	35	4376.397
	SOFA score	-TD	-2064.232	35	4198.465
	Age	-NL	-2052.281	35	4174.561
	SOFA score	-NL	-2046.978	35	4163.956
Step 3		Deletion			
	Immunosuppression	-TD	-2045.518	33	4157.036
	Cirrhosis	-TD	-2055.391	33	4174.781
	Infection type (nosocomial)	-TD	-2052.475	33	4168.950
	Age	-TD	-2153.225	32	4368.449
	SOFA score	-TD	-2070.349	32	4202.697
	Age	-NL	-2056.141	32	4174.282
	SOFA score	-NL	2048.335	32	4158.670
Step 4		Deletion			
	Cirrhosis	-TD	-2056.636	30	4173.271
	Infection type (nosocomial)	-TD	-2049.984	30	4159.969
	Age	-TD	-2153.198	29	4364.397
	SOFA score	-TD	-2070.484	29	4198.968
	Age	-NL	-2052.821	29	4163.641
	SOFA score	-NL	-2050.143	29	4158.287
Selected covariates/ Final Model	Immunosuppression (constant); Knaus score (constant); Infection site (constant); Cirrhosis (TD); Infection type (TD); Age (TD+NL); SOFA(TD+NL)				

B.4.2 Results of the estimated NL effect of age and SOFA

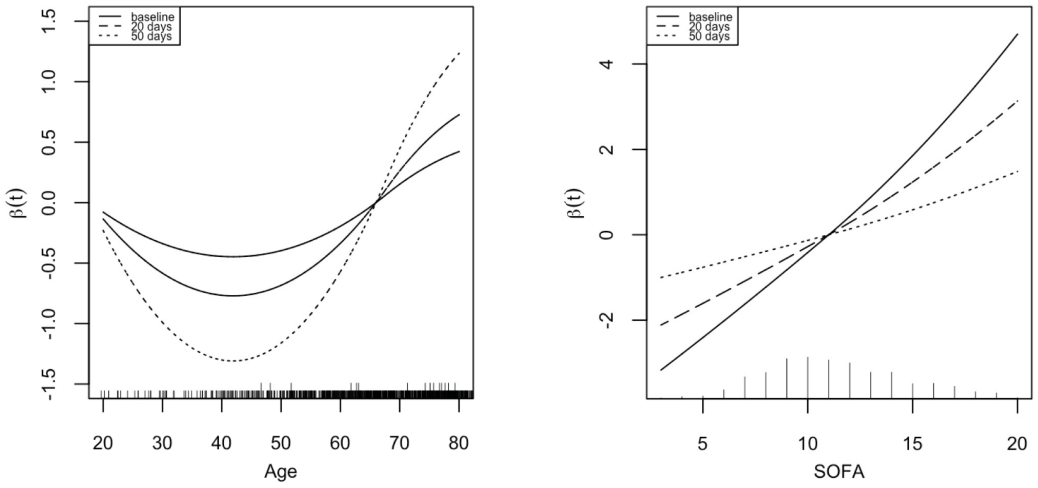


Figure B-15: Estimated NL effects of age and SOFA score, relative to the mean values, at baseline, 20 days and 50 days, respectively, from the flexible AFT model in the septic shock study.

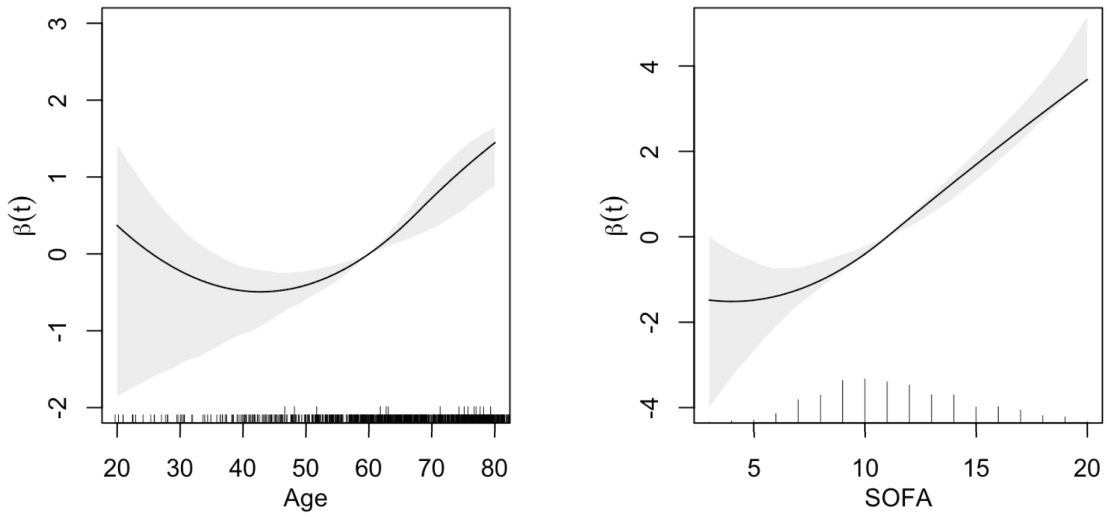


Figure B-16: Estimated NL effects of age and SOFA score, relative to the mean values, from the non-linear extended AFT model in the septic shock study.

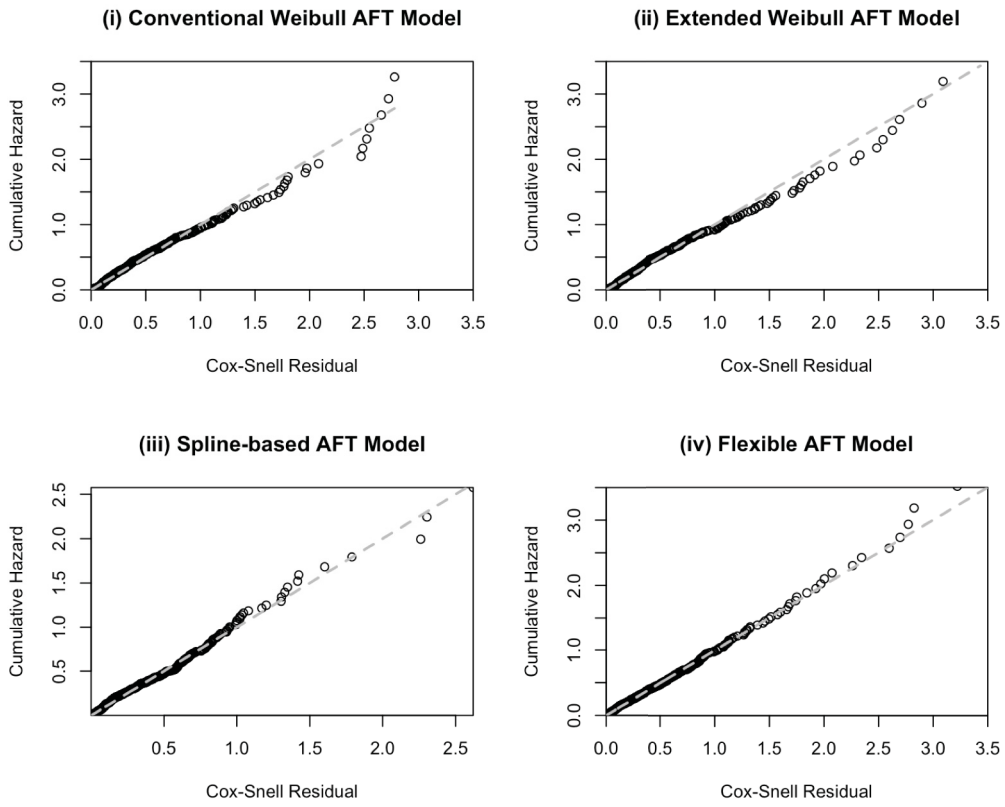


Figure B-17: Cox-Snell residual plots from alternative models in the septic shock study.

B.5 Analytical examples for TD effect and time-dependent time ratio

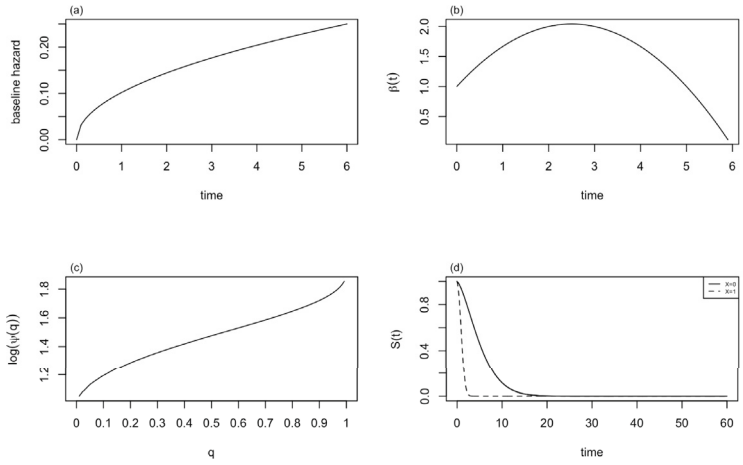


Figure B-18: Analytical example 1 that shows different patterns for TD effect $\beta(t)$ and time-dependent time ratio $\psi(q)$. (a) baseline hazard function; (b) TD effect $\beta(t)$ of a binary covariate; (c) time-dependent time ratio in the log scale $\log \psi(q)$, as a function of the quantiles of survival times, derived from inversing the survival functions in (d) for $X = 0$ vs. $X = 1$.

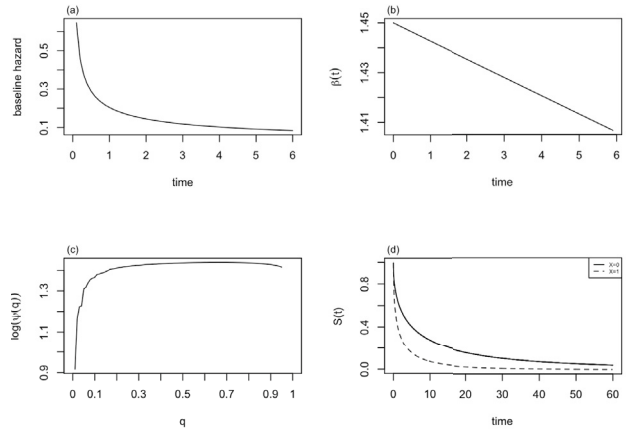


Figure B-19: Analytical example 2 that shows different patterns for TD effect $\beta(t)$ and time-dependent time ratio $\psi(q)$. (a) baseline hazard function; (b) TD effect $\beta(t)$ of a binary covariate; (c) time-dependent time ratio in the log scale $\log \psi(q)$, as a function of the quantiles of survival times, derived from inversing the survival functions in (d) for $X = 0$ vs. $X = 1$.

Appendix C: Appendix to Manuscript 3

C.1 Supplementary Materials for the non-small cell lung cancer (NSCLC) study

Table C-1: Baseline characteristics of patients in the NSCLC study (N = 269)

Variables	Descriptive statistics
Stage: IIIA/IIIB *n* (%)	70 (26.0)
IIIB+/IV *n* (%)	199 (74.0)
Smoking status: Never *n* (%)	41 (15.2)
Ever *n* (%)	228 (84.8)
Chemotherapy type: Single-agent *n* (%)	66 (24.55)
Double-agent *n* (%)	203 (75.5)
CRP	
Mean (sd)	36.2 (53.6)
Median {quartile} (range)	13.1 {4.9, 39.9} (0.3, 316.8)
$\log_2$CRP	
Mean (sd)	3.8 (2.2)
Median {quartile} (range)	3.7 {2.3, 5.3} (-1.7, 8.3)
Albumin mean (sd)	40.2 (4.1)
LDH	
Mean (sd)	248.8 (199.4)
Median {quartile} (range)	211 {169, 263} (98, 2500)
$\log_2$LDH	
Mean (sd)	7.8 (0.6)
Median {quartile} (range)	7.7 {7.4, 8.0} (6.6, 11.3)
ANC	7.09 (3.55)
Lymphocytes	1.59 (0.70)
Age: mean (sd)	64.3 (11.0)

Table C-2: Results of the backwards elimination procedure in building the final flexible TD/NL Cox PH model

	Covariates		Log-likelihood	Degree of freedom	AIC
Stage 1	10 covariates		-904.8782	58	1925.756
Selected covariates/ Initial Model	Stage(TD), Smoking(TD), Chemotherapy(TD), Sex (TD), $\log_2$CRP(TD+NL), Albumin(TD+NL), $\log_2$LDH (TD+NL), ANC (TD+NL), Lymphocytes(TD+NL), Age (TD+NL)				
Stage 2		Stage 2			
Step 1		Deletion			
	Stage	-TD	-906.5334	55	1923.067
	Smoking	-TD	-907.5901	55	1925.180
	Chemotherapy	-TD	-907.2901	55	1924.580
	Sex	-TD	-905.5291	55	1921.058
	$\log_2$CRP	-TD	905.8750	54	1919.750
	Albumin	-TD	-914.8705	54	1937.741
	$\log_2$LDH	-TD	-912.6976	54	1933.395
	ANC	-TD	-909.8859	54	1927.772
	Lymphocytes	-TD	-909.1636	54	1926.327
	Age	-TD	-911.0809	54	1930.162
	$\log_2$CRP	-NL	-906.8110	55	1923.622
	Albumin	-NL	-905.4212	55	1920.842
	$\log_2$LDH	-NL	-906.0528	55	1922.106
	ANC	-NL	-906.0443	55	1922.089
	Lymphocytes	-NL	-906.8788	55	1923.758
	Age	-NL	-911.2044	55	1932.409
Step 2		Deletion			
	Stage	-TD	-907.7092	51	1917.418
	Smoking	-TD	-909.8439	51	1921.688
	Chemotherapy	-TD	-907.9737	51	1917.947
	Sex	-TD	-906.5915	51	1915.183
	Albumin	-TD	-916.5361	50	1933.072
	$\log_2$LDH	-TD	-913.8720	50	1927.744
	ANC	-TD	-911.0939	50	1922.188
	Lymphocytes	-TD	-910.1748	50	1920.350
	Age	-TD	-913.4862	50	1926.972
	$\log_2$CRP	-NL	-907.6509	52	1919.302
	Albumin	-NL	-906.4229	51	1914.846
	$\log_2$LDH	-NL	-906.9582	51	1915.916
	ANC	-NL	-907.1764	51	1916.353
	Lymphocytes	-NL	-907.9335	51	1917.867
	Age	-NL	-912.2185	51	1926.437
Step 3		Deletion			
	Stage	-TD	-908.7579	48	1913.516
	Smoking	-TD	-910.1212	48	1916.242
	Chemotherapy	-TD	-908.5231	48	1913.046
	Sex	-TD	-907.1501	48	1910.300
	Albumin	-TD	-916.7148	48	1929.430
	$\log_2$LDH	-TD	-914.3635	47	1922.727

	ANC	-TD	-917.2551	47	1928.510
	Lymphocytes	-TD	-910.4376	47	1914.875
	Age	-TD	-913.5826	47	1921.165
	log_2CRP	-NL	-908.0399	49	1914.080
	log_2LDH	-NL	-907.7230	48	1911.446
	ANC	-NL	-907.9190	48	1911.838
	Lymphocytes	-NL	-908.2558	48	1912.512
	Age	-NL	-912.4233	48	1920.847
Step 4		Deletion			
	Stage	-TD	-910.1098	45	1910.220
	Smoking	-TD	-913.4796	45	1916.959
	Chemotherapy	-TD	-909.2689	45	1908.538
	Albumin	-TD	-917.6396	45	1925.279
	log_2LDH	-TD	-915.1598	44	1918.320
	ANC	-TD	-911.8240	44	1911.648
	Lymphocytes	-TD	-911.7530	44	1911.506
	Age	-TD	-914.1130	44	1916.226
	log_2CRP	-NL	-908.7433	46	1909.487
	log_2LDH	-NL	-908.7059	45	1907.412
	ANC	-NL	-908.7219	45	1907.444
	Lymphocytes	-NL	-909.2475	45	1908.495
	Age	-NL	-912.7979	45	1915.596
Step 5		Deletion			
	Stage	-TD	-910.5987	42	1905.197
	Smoking	-TD	-914.5138	42	1913.028
	Chemotherapy	-TD	-910.2454	42	1904.491
	Albumin	-TD	-918.1411	42	1920.282
	log_2LDH	-TD	-915.2628	42	1914.526
	ANC	-TD	-912.7823	41	1907.565
	Lymphocytes	-TD	-912.2793	41	1906.559
	Age	-TD	-914.5741	41	1911.148
	log_2CRP	-NL	-909.5354	43	1905.071
	ANC	-NL	-909.9271	42	1903.854
	Lymphocytes	-NL	-910.1732	42	1904.346
	Age	-NL	-911.1712	42	1906.342
Step 6		Deletion			
	Stage	-TD	-912.2870	39	1902.574
	Smoking	-TD	-916.5047	39	1911.009
	Chemotherapy	-TD	-912.2094	39	1902.419
	Albumin	-TD	-919.4016	39	1916.803
	log_2LDH	-TD	-916.7141	39	1911.428
	ANC	-TD	-913.5680	39	1905.136
	Lymphocytes	-TD	-914.4134	38	1904.827
	Age	-TD	-916.1569	38	1908.314
	log_2CRP	-NL	-911.6018	40	1903.204
	Lymphocytes	-NL	-912.5404	39	1903.081
	Age	-NL	-913.0323	39	1904.065
Step 7		Deletion			
	Stage	-TD	-914.6833	36	1901.367
	Smoking	-TD	-918.7299	36	1909.460

	Albumin	-TD	-923.1595	36	1918.319
	log_2LDH	-TD	-919.4506	36	1910.901
	ANC	-TD	-915.6525	36	1903.305
	Lymphocytes	-TD	-916.8842	35	1903.768
	Age	-TD	-918.1633	35	1906.327
	log_2CRP	-NL	-913.9551	37	1901.910
	Lymphocytes	-NL	-915.1749	36	1902.350
	Age	-NL	-915.0727	36	1902.145
Step 8		Deletion			
	Smoking	-TD	-921.6972	33	1909.394
	Albumin	-TD	-924.6746	33	1915.349
	log_2LDH	-TD	-920.4553	33	1906.911
	ANC	-TD	-918.278	33	1902.556
	Lymphocytes	-TD	-917.3820	32	1898.764
	Age	-TD	-919.4050	32	1902.810
	log_2CRP	-NL	-915.9682	34	1899.936
	Lymphocytes	-NL	-915.9785	33	1897.957
	Age	-NL	-916.4058	33	1898.812
Step 9		Deletion			
	Smoking	-TD	-922.4666	30	1904.933
	Albumin	-TD	-926.2143	30	1912.429
	log_2LDH	-TD	-921.4425	30	1902.885
	ANC	-TD	-920.6660	30	1901.332
	Lymphocytes	-TD	-918.0590	30	1896.118
	Age	-TD	-920.4081	29	1898.816
	log_2CRP	-NL	-918.2958	31	1898.592
	Age	-NL	-919.8491	30	1899.698
Step 10		Deletion			
	Smoking	-TD	-923.3304	27	1900.661
	Albumin	-TD	-927.9397	27	1909.879
	log_2LDH	-TD	-922.9626	27	1899.925
	ANC	-TD	-922.5874	27	1899.175
	Age	-TD	-921.1334	26	1894.267
	log_2CRP	-NL	-920.4000	28	1896.800
	Age	-NL	-920.3759	27	1894.752
Step 11		Deletion			
	Smoking	-TD	-924.0146	23	1894.029
	Albumin	-TD	-931.5921	23	1909.184
	log_2LDH	-TD	-926.1222	23	1898.244
	ANC	-TD	-924.6809	23	1895.362
	log_2CRP	-NL	-922.7632	24	1893.526
	Age	-NL	-922.1241	24	1892.248
Step 12		Deletion			
	Smoking	-TD	-924.9843	21	1891.969
	Albumin	-TD	-932.8248	21	1907.650
	log_2LDH	-TD	-926.7726	21	1895.545
	ANC	-TD	-925.6975	21	1893.395
	log_2CRP	-NL	-924.6073	22	1893.215
	Albumin	-TD	-935.2750	18	1906.550
	log_2LDH	-TD	-928.8652	18	1893.730

	ANC	-TD	-929.1288	18	1894.258
	log_2CRP	-NL	-927.3040	19	1892.608
Selected covariates/ Final Model	Stage(constant), Smoking (constant), Chemotherapy(constant), sex (constant), log_2CRP(NL), Albumin(TD), log_2LDH (TD), ANC (TD), Lymphocytes(PH+LL), Age (PH+LL)				

Abbreviations: CRP=C-reactive protein; LDH=lactate dehydrogenase; ANC=Absolute neutrophil counts; AIC=Akaike information criterion

Table C-3: Results of the backwards elimination procedure in building the final flexible TD/NL AFT model

Covariates			Log-likelihood	Degree of freedom	AIC
Stage 1	10 covariates		-665.0755	70	1470.151
Selected covariates/ Initial Model	Stage (TD), Smoking (TD), Chemotherpay (TD), Sex (TD), $\log_2$CRP(TD+NL), Albumin(TD+NL), $\log_2$LDH (TD+NL), ANC (TD+NL), Lymphocytes(TD+NL), Age (TD+NL)				
Stage 2		Stage 2			
Step 1		Deletion			
	Stage	-TD	-662.4757	67	1458.951
	Smoking	-TD	-665.7818	67	1465.564
	Chemotherapy	-TD	-668.3102	67	1470.620
	Sex	-TD	-664.7320	67	1463.464
	$\log_2$CRP	-TD	-664.7395	66	1461.479
	Albumin	-TD	-686.7901	66	1505.580
	$\log_2$LDH	-TD	-678.0230	66	1488.046
	ANC	-TD	-704.5850	66	1541.170
	Lymphocytes	-TD	-694.1479	66	1520.296
	Age	-TD	-662.5518	66	1457.104
	$\log_2$CRP	-NL	-665.4292	66	1462.858
	Albumin	-NL	-662.9892	66	1457.978
	$\log_2$LDH	-NL	-668.6128	66	1469.226
	ANC	-NL	-665.1000	66	1462.200
	Lymphocytes	-NL	-666.3311	66	1464.662
	Age	-NL	-666.9560	66	1465.912
Step 2		Deletion			
	Stage	-TD	-700.3301	63	1526.660
	Smoking	-TD	-667.9859	63	1461.972
	Chemotherapy	-TD	-678.0407	63	1482.081
	Sex	-TD	-666.7105	63	1459.421
	$\log_2$CRP	-TD	-772.5718	62	1669.144
	Albumin	-TD	-881.1686	62	1886.337
	$\log_2$LDH	-TD	-693.7866	62	1511.573
	ANC	-TD	-766.3042	62	1656.608
	Lymphocytes	-TD	-682.9132	62	1489.826
	$\log_2$CRP	-NL	-672.5273	62	1469.055
	Albumin	-NL	-667.7165	62	1459.433
	$\log_2$LDH	-NL	-688.7646	62	1501.529
	ANC	-NL	-657.5223	62	1439.045
	Lymphocytes	-NL	-666.9404	62	1457.881
	Age	-NL	-668.1827	63	1462.365
Step 3		Deletion			
	Stage	-TD	-682.8068	59	1483.614
	Smoking	-TD	-676.7660	59	1471.532
	Chemotherapy	-TD	-667.0727	59	1452.145
	Sex	-TD	-671.3725	59	1460.745
	$\log_2$CRP	-TD	-743.6007	58	1603.201
	Albumin	-TD	-677.8076	58	1471.615

	log_2LDH	-TD	-672.6906	58	1461.381
	ANC	-TD	-668.7434	59	1455.487
	Lymphocytes	-TD	-675.8479	58	1467.696
	log_2CRP	-NL	-704.8040	58	1525.608
	Albumin	-NL	-660.3191	58	1436.638
	log_2LDH	-NL	-677.0087	58	1470.017
	Lymphocytes	-NL	-663.0342	58	1442.068
	Age	-NL	-661.3590	59	1440.718
Step 4		Deletion			
	Stage	TD	-662.0972	55	1434.194
	Smoking	TD	-663.7005	55	1437.401
	Chemotherapy	TD	-667.0983	55	1444.197
	Sex	TD	-665.7052	55	1441.410
	log_2CRP	TD	-673.4709	54	1454.942
	Albumin	TD	-680.8528	55	1471.706
	log_2LDH	TD	-679.8092	54	1467.618
	ANC	TD	-682.0629	55	1474.126
	Lymphocytes	TD	-680.6855	54	1469.371
	log_2CRP	NL	-676.3803	54	1460.761
	log_2LDH	NL	-689.6708	54	1487.342
	Lymphocytes	NL	-664.7590	54	1437.518
	Age	NL	-668.6763	55	1447.353
Step 5		Deletion			
	Smoking	TD	-665.4570	52	1434.914
	Chemotherapy	TD	-665.3920	52	1434.784
	Sex	TD	-665.4737	52	1434.947
	log_2CRP	TD	-688.9909	51	1479.982
	Albumin	TD	-675.4861	52	1454.972
	log_2LDH	TD	-681.8193	51	1465.639
	ANC	TD	-668.1680	52	1440.336
	Lymphocytes	TD	-664.9544	51	1431.909
	log_2CRP	NL	-665.8842	51	1433.768
	log_2LDH	NL	-678.9901	51	1459.980
	Lymphocytes	NL	-665.3980	51	1432.796
	Age	NL	-681.2766	52	1466.553
Step 6		Deletion			
	Smoking	TD	-666.0313	48	1428.063
	Chemotherapy	TD	-684.9225	48	1465.845
	Sex	TD	-678.2013	48	1452.403
	log_2CRP	TD	-669.7692	47	1433.538
	Albumin	TD	-670.8707	48	1437.741
	log_2LDH	TD	-682.0652	47	1458.130
	ANC	TD	-677.7420	48	1451.484
	log_2CRP	NL	-675.0278	47	1444.056
	log_2LDH	NL	-668.0166	47	1430.033
	Lymphocytes	NL	-665.9606	48	1427.921
	Age	NL	-666.7439	48	1429.488
Step 7		Deletion			
	Smoking	TD	-675.5571	45	1441.114
	Chemotherapy	TD	-668.6806	45	1427.361

	Sex	TD	-666.8228	45	1423.645
	log_2CRP	TD	-686.8774	44	1461.755
	Albumin	TD	-678.2320	45	1446.464
	log_2LDH	TD	-678.2117	44	1444.423
	ANC	TD	-680.1667	45	1450.333
	log_2CRP	NL	-676.0199	44	1440.040
	log_2LDH	NL	-669.4303	44	1426.861
	Age	NL	-667.9330	45	1425.866
Step 8	Deletion				
	Smoking	TD	-719.9066	42	1523.813
	Chemotherapy	TD	-669.7488	42	1423.498
	log_2CRP	TD	-680.8230	41	1443.646
	Albumin	TD	-685.9158	42	1455.832
	log_2LDH	TD	-678.1219	41	1438.244
	ANC	TD	-693.8110	42	1471.622
	log_2CRP	NL	-675.4547	41	1432.909
	log_2LDH	NL	-669.6709	41	1421.342
	Age	NL	-670.3024	42	1424.605
Step 9	Deletion				
	Smoking	TD	-671.4910	38	1418.982
	Chemotherapy	TD	-671.8274	38	1419.655
	log_2CRP	TD	-734.3446	37	1542.689
	Albumin	TD	-678.4693	38	1432.939
	log_2LDH	TD	-703.4496	38	1482.899
	ANC	TD	-788.3677	38	1652.735
	log_2CRP	NL	-671.9115	37	1417.823
	Age	NL	-672.1038	38	1420.208
Step 10	Deletion				
	Smoking	TD	-673.1075	34	1414.215
	Chemotherapy	TD	-674.006	34	1416.012
	log_2CRP	TD	-730.0440	34	1528.088
	Albumin	TD	-680.2790	34	1428.558
	log_2LDH	TD	-878.0193	34	1824.039
	ANC	TD	-828.8082	34	1725.616
	Age	NL	-676.8697	34	1421.739
Step 11	Chemotherapy	TD	-695.3668	31	1452.733
	log_2CRP	TD	-680.8642	31	1423.728
	Albumin	TD	-680.5861	31	1423.172
	log_2LDH	TD	-703.5252	31	1469.050
	ANC	TD	-745.0766	31	1552.153
	Age	NL	-686.4954	31	1434.991
Selected covariates/ Final Model	Stage (constant), Smoking (constant), Chemotherapy (TD), sex (constant), log_2CRP(TD), Albumin(TD), log_2LDH (TD), ANC (TD), Lymphocytes(constant+LL), Age (NL)				

Abbreviations: CRP=C-reactive protein; LDH=lactate dehydrogenase; ANC=Absolute neutrophil counts; AIC=Akaike information criterion

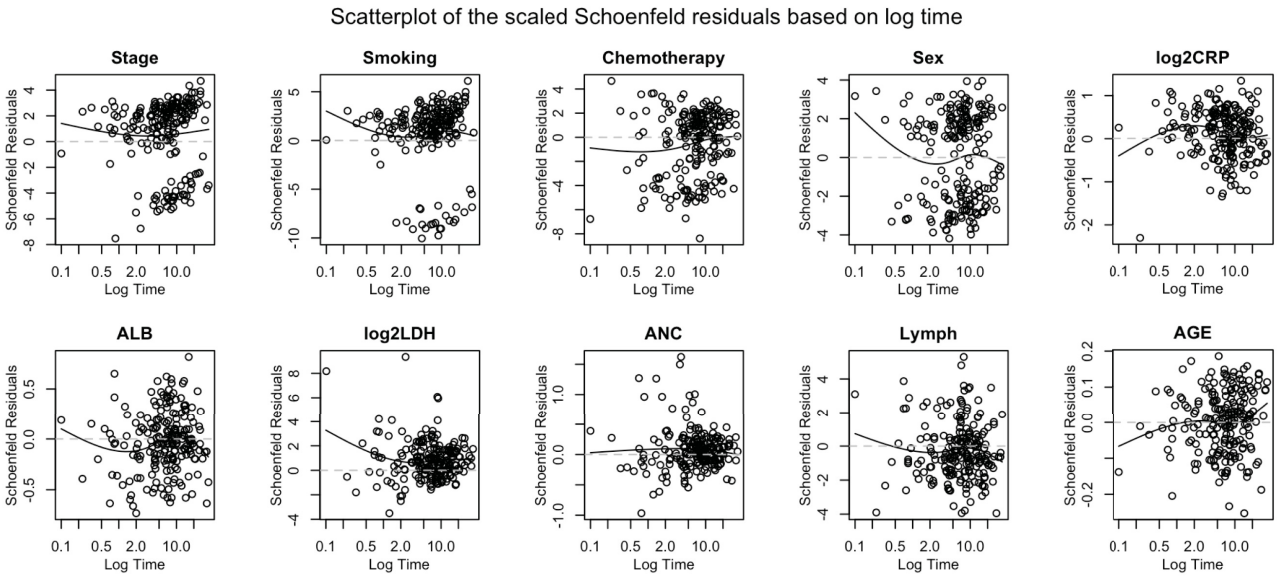


Figure C-1: Schoenfeld residual plots for checking the Cox PH assumption

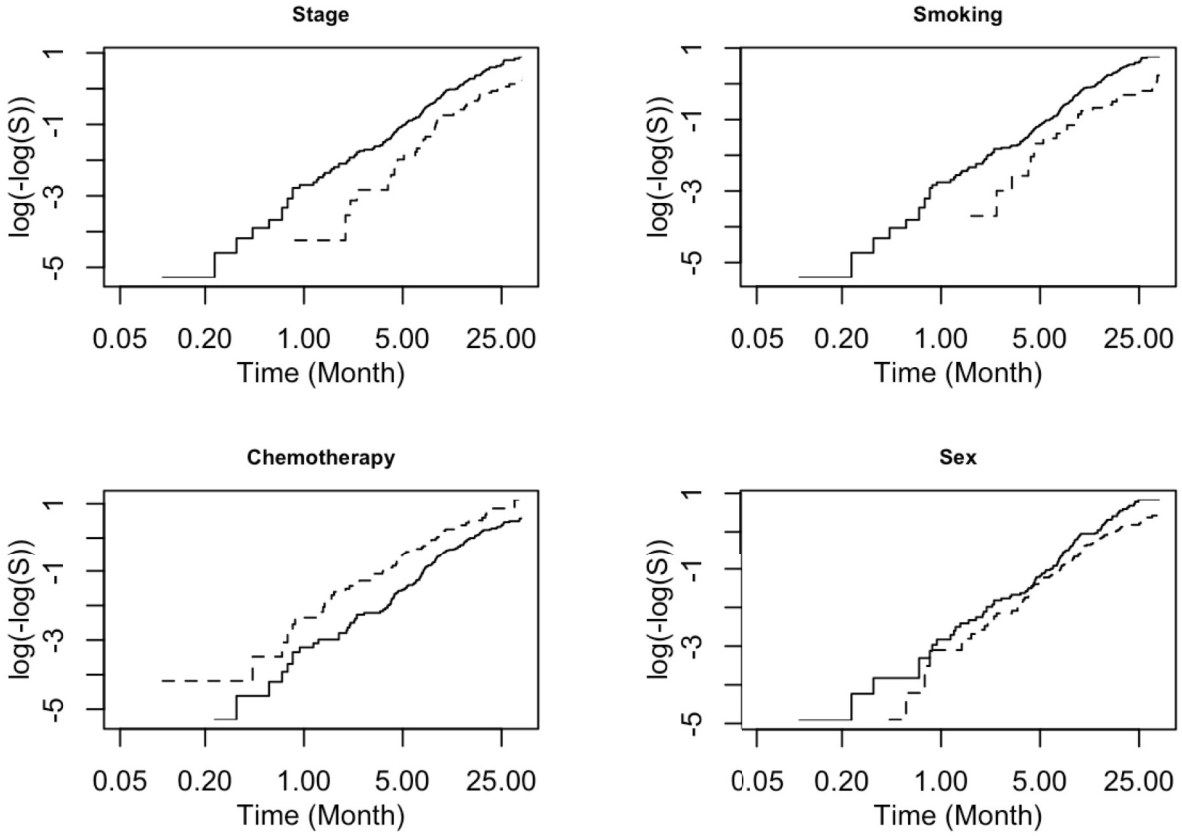


Figure C-2: Log-log curves for checking the Cox PH assumption

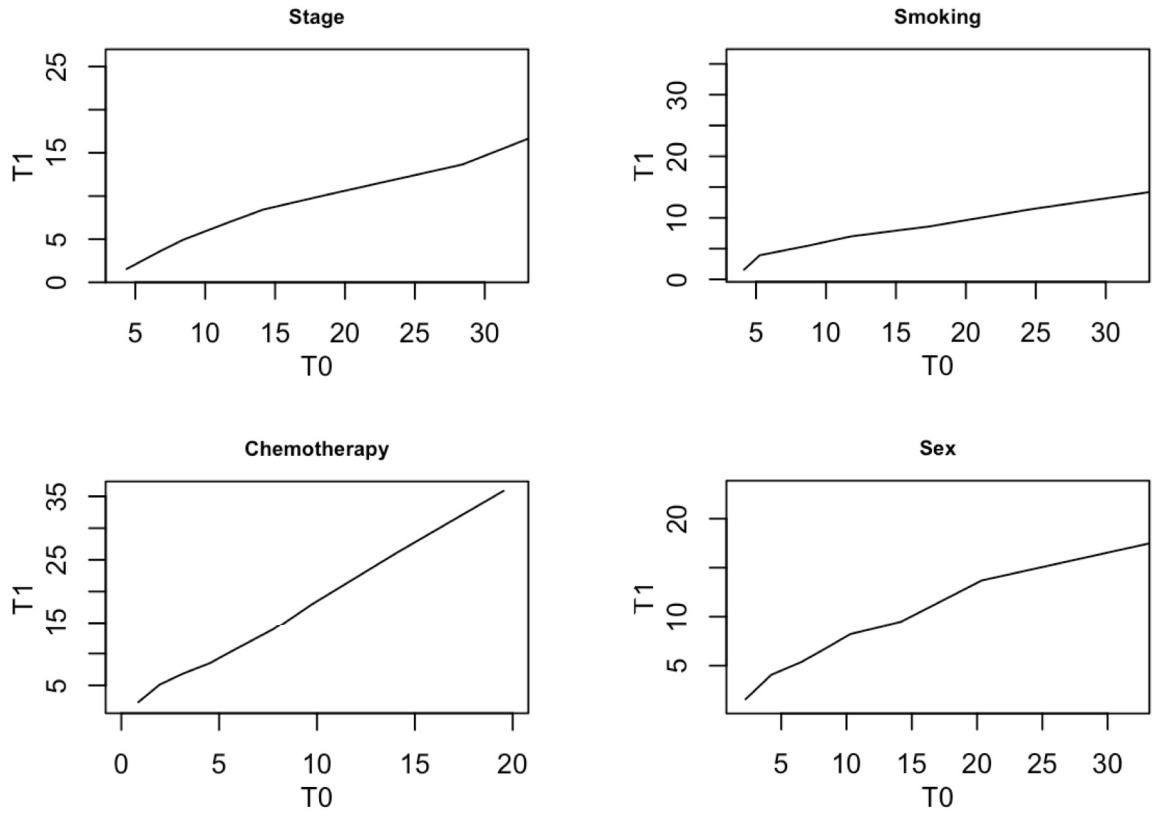


Figure C-3: QQ plots for checking the AFT constant time ratio assumption

The calculation of the full log-likelihood for the PH model (i)

We reply on a spline-based PH model to obtain the full log-likelihood of model (i). This model extends the conventional Cox PH model to model the baseline hazard function with regression B-spline. Therefore, it can be considered as a special case of the flexible TD/NL Cox PH model (6.2.1) with *a priori* linearity and constant effects. The full log-likelihood is calculated by two estimation steps. In the first step, the hazard ratios are estimated by maximizing the partial likelihood of the conventional Cox PH model. Then, in the second step, we plug the estimated hazard ratios into the spline-based full likelihood, which is maximized in order to estimate the baseline hazard function. The maximized full log-likelihood is reported in Table 6-1 to facilitate the comparison between the conventional Cox PH (i) and the other

alternative models (ii-vi). The resulting estimated hazard and survival curve for the hypothetical reference group are reported in Figure 6-4.

REFERENCES

[1] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[2] Nils Lid Hjort. On inference in parametric survival data models. *International Statistical Review/Revue Internationale de Statistique*, pages 355–387, 1992.

[3] David Roxbee Cox and David Oakes. *Analysis of survival data*, volume 21. CRC Press, 1984.

[4] William J Mackillop and Carol F Quirt. Measuring the accuracy of prognostic judgments in oncology. *Journal of clinical epidemiology*, 50(1):21–29, 1997.

[5] Marilyse Julien and James A Hanley. Profile-specific survival estimates: making reports of clinical trials more patient-relevant. *Clinical Trials*, 5(2):107–115, 2008.

[6] Lee-Jen Wei. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in medicine*, 11(14-15):1871–1879, 1992.

[7] John D Kalbfleisch and Ross L Prentice. *The statistical analysis of failure time data*, volume 360. John Wiley & Sons, 2011.

[8] Hajime Uno, Janet Wittes, Haoda Fu, Scott D Solomon, Brian Claggett, Lu Tian, Tianxi Cai, Marc A Pfeffer, Scott R Evans, and Lee-Jen Wei. Alternatives to hazard ratios for comparing the efficacy or safety of therapies in noninferiority studies. *Annals of internal medicine*, 163(2):127–134, 2015.

[9] Jesus Orbe, Eva Ferreira, and Vicente Núñez-Antón. Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in medicine*, 21(22):3493–3510, 2002.

[10] Jane L Hutton and PF Monaghan. Choice of parametric accelerated life and proportional hazards models for survival data: asymptotic results. *Lifetime data analysis*, 8(4):375–393, 2002.

[11] Yi-Hwei Li, John P Klein, and ML Moeschberger. Effects of model misspecification in estimating covariate effects in survival analysis for small sample sizes. *Computational statistics & data analysis*, 22(2):177–192, 1996.

[12] Jonathan Buckley and Ian James. Linear regression with censored data. *Biometrika*, 66(3):429–436, 1979.

[13] Tze Leung Lai and Zhiliang Ying. Large sample theory of a modified buckley-james estimator for regression analysis with censored data. *The Annals of Statistics*, pages 1370–1402, 1991.

[14] Yaacov Ritov. Estimation in a linear regression model with censored data. *The Annals of Statistics*, pages 303–328, 1990.

[15] Michael P Jones. A class of semiparametric regressions for the accelerated failure time model. *Biometrika*, 84(1):73–84, 1997.

[16] Anastasios A Tsiatis. Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics*, pages 354–372, 1990.

[17] Lee-Jen Wei, Zhiliang Ying, and DY Lin. Linear regression analysis of censored survival data based on rank tests. *Biometrika*, 77(4):845–851, 1990.

[18] Donglin Zeng and DY Lin. Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association*, 102(480):1387–1396, 2007.

[19] Zhezhen Jin, DY Lin, LJ Wei, and Zhiliang Ying. Rank-based inference for the accelerated failure time model. *Biometrika*, 90(2):341–353, 2003.

[20] Jian Huang, Shuangge Ma, and Huiliang Xie. Least absolute deviations estimation for the accelerated failure time model. *Statistica Sinica*, pages 1533–1548, 2007.

[21] Jeremy MG Taylor. Semi-parametric estimation in failure time mixture models. *Biometrics*, pages 899–907, 1995.

[22] Chin-Shang Li and Jeremy MG Taylor. A semi-parametric accelerated failure time cure model. *Statistics in medicine*, 21(21):3235–3247, 2002.

[23] Jiajia Zhang and Yingwei Peng. A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statistics in medicine*, 26(16):3157–3171, 2007.

[24] A Ciampi and J Etezadi-Amoli. A general model for testing the proportional hazards and the accelerated failure time hypotheses in the analysis of censored survival data with covariates. *Communications in Statistics-Theory and Methods*, 14(3):651–667, 1985.

[25] Jamshid Etezadi-Amoli and Antonio Ciampi. Extended hazard regression for censored survival data with covariates: a spline approximation for the baseline hazard function. *Biometrics*, pages 181–192, 1987.

[26] Arnošt Komárek, Emmanuel Lesaffre, and Joan F Hilton. Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *Journal of Computational and Graphical Statistics*, 14(3):726–745, 2005.

[27] Jesus Orbe, Eva Ferreira, and Vicente Núñez-Antón. Censored partial regression. *Biostatistics*, 4(1):109–121, 2003.

[28] TA Gerds and Martin Schumacher. On functional misspecification of covariates in the cox regression model. *Biometrika*, 88(2):572–580, 2001.

[29] Miguel A Hernán. The hazards of hazard ratios. *Epidemiology (Cambridge, Mass.)*, 21(1):13, 2010.

[30] Stephen W Lagakos. General right censoring and its impact on the analysis of survival data. *Biometrics*, pages 139–156, 1979.

[31] Michal Abrahamowicz, Roxane du Berger, and Steven A Graver. Flexible modeling of the effects of serum cholesterol on coronary heart disease mortality. *American journal of epidemiology*, 145(8):714–729, 1997.

[32] Dexiang Gao, Gary K Grunwald, John S Rumsfeld, Lynn Schooley, Todd MacKenzie, and A Laurie W Shroyer. Time-varying risk factors for long-term mortality after coronary artery bypass graft surgery. *The Annals of thoracic surgery*, 81(3):793–799, 2006.

[33] Willy Wynant and Michal Abrahamowicz. Validation of the alternating conditional estimation algorithm for estimation of flexible extensions of cox's proportional hazards model with nonlinear constraints on the parameters. *Biometrical Journal*, 58(6):1445–1464, 2016.

[34] Willy Wynant and Michal Abrahamowicz. Impact of the model-building strategy on inference about nonlinear and time-dependent covariate effects in survival analysis. *Statistics in medicine*, 33(19):3318–3337, 2014.

[35] B Gagnon, M Abrahamowicz, Y Xiao, ME Beauchamp, N MacDonald, G Kasymjanova, H Kreisman, and D Small. Flexible modeling improves assessment of prognostic value of c-reactive protein in advanced non-small cell lung cancer. *British journal of cancer*, 102(7):1113–1122, 2010.

[36] Lynn A Sleeper and David P Harrington. Regression splines in the cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association*, 85(412):941–949, 1990.

[37] Robert J Gray. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, 87(420):942–951, 1992.

[38] C Cluze, Marc Colonna, Laurent Remontet, Florence Poncet, E Sellier, Arnaud Seigneurin, Patricia Delafosse, and Nadine Bossard. Analysis of the effect of age on the prognosis of breast cancer. *Breast cancer research and treatment*, 117(1):121, 2009.

[39] Willi Sauerbrei, Patrick Royston, and Maxime Look. A new proposal for multivariable modelling of time-varying effects in survival data based on fractional polynomial time-transformation. *Biometrical Journal*, 49(3):453–473, 2007.

[40] Michal Abrahamowicz and Todd A MacKenzie. Joint estimation of time-dependent and non-linear effects of continuous covariates on survival. *Statistics in medicine*, 26(2):392–408, 2007.

[41] Willy Wynant and Michal Abrahamowicz. Flexible estimation of survival curves conditional on non-linear and time-dependent predictor effects. *Statistics in medicine*, 35(4):553–565, 2016.

[42] Yubo Zou, Jiajia Zhang, and Guoyou Qin. A semiparametric accelerated failure time partial linear model and its application to breast cancer. *Computational statistics & data analysis*, 55(3):1479–1487, 2011.

[43] Hongqi Xue, KF Lam, Benjamin J Cowling, and Frank de Wolf. Semi-parametric accelerated failure time regression analysis with application to interval-censored hiv/aids data. *Statistics in medicine*, 25(22):3850–3863, 2006.

[44] Elsayed A Elsayed, Haitao Liao, and Xindong Wang. An extended linear hazard regression model with application to time-dependent dielectric breakdown of thermal oxides. *Iie Transactions*, 38(4):329–340, 2006.

[45] Catherine Quantin, Michal Abrahamowicz, Thierry Moreau, Gillian Bartlett, Todd MacKenzie, Mohammed Adnane Tazi, Luc Lalonde, and Jean Faivre. Variation over time of the effects of prognostic factors in a population-based study of colon cancer: comparison of statistical models. *American journal of epidemiology*, 150(11):1188–1200, 1999.

[46] L Remontet, N Bossard, A Belot, and J FRANCIM Esteve. An overall strategy based on regression models to estimate relative survival and models to estimate relative survival and model the effects of prognostic factors in cancer survival studies. *Stat Med*, 26(10):2214–2228, 2007.

[47] Adam C Salisbury, Kimberly J Reid, and John A Spertus. Impact of chronic obstructive pulmonary disease on post-myocardial infarction outcomes. *The American journal of cardiology*, 99(5):636–641, 2007.

[48] Soudabeh Khojasteh Banankhah, Erika Friedmann, and Sue Thomas. Effective treatment of depression improves post-myocardial infarction survival. *World journal of cardiology*, 7(4):215, 2015.

[49] John N Primrose, Rafael Perera, Alastair Gray, Peter Rose, Alice Fuller, Andrea Corkhill, Steve George, and David Mant. Effect of 3 to 5 years of scheduled cea and ct follow-up to detect recurrence of colorectal cancer: the facs randomized clinical trial. *Jama*, 311(3):263–270, 2014.

[50] Harry Bartelink, Jean-Claude Horiot, Philip Poortmans, Henk Struikmans, Walter Van den Bogaert, Isabelle Barillot, Alain Fourquet, Jacques Borger, Jos Jager, Willem Hoogenraad, et al. Recurrence rates after treatment of breast cancer with standard radiotherapy with or without additional radiation. *New England Journal of Medicine*, 345(19):1378–1387, 2001.

[51] Mario Cleves, William Gould, William W Gould, Roberto Gutierrez, and Yulia Marchenko. *An introduction to survival analysis using Stata*. Stata press, 2008.

[52] John P Klein and Melvin L Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2006.

[53] David G Kleinbaum and Mitchel Klein. *Survival analysis*. Springer, 2010.

[54] Jerald F Lawless. *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons, 2011.

[55] Jay D Teachman and Mark D Hayward. Interpreting hazard rate models. *Sociological methods & research*, 21(3):340–371, 1993.

[56] NE Breslow. Discussion of the paper by dr cox. *Journal of the Royal Statistical Society, Series B*, 34:216–217, 1972.

[57] DY Lin. On the breslow estimator. *Lifetime data analysis*, 13(4):471–480, 2007.

[58] Terry M Therneau, Patricia M Grambsch, and Thomas R Fleming. Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160, 1990.

[59] Kenneth R Hess. Graphical methods for assessing violations of the proportional hazards assumption in cox regression. *Statistics in medicine*, 14(15):1707–1723, 1995.

[60] David Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241, 1982.

[61] Frank E Harrell Jr. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer, 2015.

[62] Nicholas H Ng'andu. An empirical comparison of statistical tests for assessing the proportional hazards assumption of cox's model. *Statistics in medicine*, 16(6):611–626, 1997.

[63] DG Altman, BL De Stavola, SB Love, and KA Stepniewska. Review of survival analyses published in cancer journals. *British journal of cancer*, 72(2):511–518, 1995.

[64] Miguel A Hernán, Sonia Hernández-Díaz, and James M Robins. A structural approach to selection bias. *Epidemiology*, pages 615–625, 2004.

[65] W Dana Flanders and Mitchel Klein. Properties of 2 counterfactual effect definitions of a point exposure. *Epidemiology*, pages 453–460, 2007.

[66] Mitchell H Gail, S Wieand, and Steven Piantadosi. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, 71(3):431–444, 1984.

[67] Cyntha A Struthers and John D Kalbfleisch. Misspecified proportional hazard models. *Biometrika*, 73(2):363–369, 1986.

[68] Jean Bretagnolle and Catherine Huber-Carol. Effects of omitting covariates in cox's model for survival data. *Scandinavian journal of statistics*, pages 125–138, 1988.

[69] Claudia Schmoor and Martin Schumacher. Effects of covariate omission and categorization when analysing randomized trials with the cox model. *Statistics in medicine*, 16(3):225–237, 1997.

[70] Thomas R Fleming and DY Lin. Survival analysis in clinical trials: past developments and future directions. *Biometrics*, 56(4):971–983, 2000.

[71] Terry Therneau. A package for survival analysis in s. version 2.38, 2015.

[72] Brandon George, Samantha Seals, and Inmaculada Aban. Survival analysis and regression models. *Journal of nuclear cardiology*, 21(4):686–694, 2014.

[73] William R Swindell. Accelerated failure time models provide a useful statistical framework for aging research. *Experimental gerontology*, 44(3):190–200, 2009.

[74] Shankar Prasad Khanal, V Sreenivas, and Subrat K Acharya. Accelerated failure time models: an application in the survival of acute liver failure patients in india. *Int J Sci Res*, 3:161–66, 2014.

[75] Nancy Reid. A conversation with sir david cox. *Statistical Science*, 9(3):439–455, 1994.

[76] Sander Greenland. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology*, pages 356–365, 1995.

[77] Patrick Royston and Willi Sauerbrei. *Multivariable model-building: a pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables*, volume 777. John Wiley & Sons, 2008.

[78] Willi Sauerbrei, Michal Abrahamowicz, Douglas G Altman, Saskia le Cessie, James Carpenter, and STRATOS initiative. Strengthening analytical thinking for observational studies: the stratos initiative. *Statistics in medicine*, 33(30):5413–5432, 2014.

[79] Hermann Brenner and Maria Blettner. Controlling for continuous confounders in epidemiologic research. *Epidemiology*, pages 429–434, 1997.

[80] Lue Ping Zhao and Laurence N Kolonel. Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies. *American journal of epidemiology*, 136(4):464–474, 1992.

[81] Trevor J Hastie. Generalized additive models. In *Statistical models in S*, pages 249–307. Routledge, 2017.

[82] Patrick Royston and Douglas G Altman. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(3):429–453, 1994.

[83] Andrea Benedetti and Michal Abrahamowicz. Using generalized additive models to reduce residual confounding. *Statistics in medicine*, 23(24):3781–3801, 2004.

[84] Paul Speckman. Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(3):413–436, 1988.

[85] David R Cox and E Joyce Snell. A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):248–265, 1968.

[86] Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pages 199–213. Springer, 1998.

[87] Chris T Volinsky and Adrian E Raftery. Bayesian information criterion for censored survival models. *Biometrics*, 56(1):256–262, 2000.

[88] Scott I Vrieze. Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods*, 17(2):228, 2012.

[89] Kenneth P Burnham and David R Anderson. Practical use of the information-theoretic approach. In *Model selection and inference*, pages 75–117. Springer, 1998.

[90] Yuhyun Park and LJ Wei. Estimating subject-specific survival functions under the accelerated failure time model. *Biometrika*, 90(3):717–723, 2003.

[91] Komárek. smoothsurv: Survival regression with smoothed error distribution, 2018.

[92] Sylvain Durrleman and Richard Simon. Flexible regression models with cubic splines. *Statistics in medicine*, 8(5):551–561, 1989.

[93] James O Ramsay et al. Monotone regression splines in action. *Statistical science*, 3(4):425–441, 1988.

[94] Carl De Boor, Carl De Boor, Etats-Unis Mathématicien, Carl De Boor, and Carl De Boor. *A practical guide to splines*, volume 27. springer-verlag New York, 1978.

[95] Michal Abrahamowicz, Todd Mackenzie, and John M Esdaile. Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. *Journal of the American Statistical Association*, 91(436):1432–1439, 1996.

[96] Christine Binquet, Michal Abrahamowicz, Karine Astruc, Jean Faivre, Claire Bonithon-Kopp, and Catherine Quantin. Flexible statistical models provided new insights into the role of quantitative prognostic factors for mortality in gastric cancer. *Journal of clinical epidemiology*, 62(3):232–240, 2009.

[97] Amel Mahboubi, Michal Abrahamowicz, Roch Giorgi, Christine Binquet, Claire Bonithon-Kopp, and Catherine Quantin. Flexible modeling of the effects of continuous prognostic factors in relative survival. *Statistics in medicine*, 30(12):1351–1365, 2011.

[98] Kenneth R Hess. Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Statistics in medicine*, 13(10):1045–1062, 1994.

[99] Philip S Rosenberg, Hormuzd Katki, Christine A Swanson, Linda M Brown, Sholom Wacholder, and Robert N Hoover. Quantifying epidemiologic risk factors using non-parametric regression: model selection remains the greatest challenge. *Statistics in medicine*, 22(21):3369–3381, 2003.

[100] Michael Hauptmann, Jürgen Wellmann, Jay H Lubin, Philip S Rosenberg, and Lothar Kreienbrock. Analysis of exposure-time-response relationships using a spline weight function. *Biometrics*, 56(4):1105–1108, 2000.

[101] Charles Kooperberg, Charles J Stone, and Young K Truong. Hazard regression. *Journal of the American Statistical Association*, 90(429):78–94, 1995.

[102] Mary J Lindstrom. Penalized estimation of free-knot splines. *Journal of Computational and Graphical Statistics*, 8(2):333–352, 1999.

[103] Patrick Royston and Mahesh KB Parmar. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in medicine*, 21(15):2175–2197, 2002.

[104] Patrick Royston, Paul C Lambert, et al. *Flexible parametric survival analysis using Stata: beyond the Cox model*. Stata Press College Station, TX, 2011.

[105] Erika Graf, Claudia Schmoor, Willi Sauerbrei, and Martin Schumacher. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*, 18(17-18):2529–2545, 1999.

[106] Thomas A Gerds and Martin Schumacher. Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biometrical Journal*, 48(6):1029–1040, 2006.

[107] Nico JD Nagelkerke et al. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692, 1991.

[108] M Schemper. The explained variation in proportional hazards regression. *Biometrika*, 77(1):216–218, 1990.

[109] Michael Schemper. The relative importance of prognostic factors in studies of survival. *Statistics in medicine*, 12(24):2377–2382, 1993.

[110] Frank E Harrell, Robert M Califf, David B Pryor, Kerry L Lee, and Robert A Rosati. Evaluating the yield of medical tests. *Jama*, 247(18):2543–2546, 1982.

[111] Frank E Harrell Jr, Kerry L Lee, Robert M Califf, David B Pryor, and Robert A Rosati. Regression modelling strategies for improved prognostic prediction. *Statistics in medicine*, 3(2):143–152, 1984.

[112] Hajime Uno, Tianxi Cai, Michael J Pencina, Ralph B D'Agostino, and LJ Wei. On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine*, 30(10):1105–1117, 2011.

[113] Hajime Uno, Tianxi Cai, Lu Tian, and Lee-Jen Wei. Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association*, 102(478):527–537, 2007.

[114] Patrick J Heagerty, Thomas Lumley, and Margaret S Pepe. Time-dependent roc curves for censored survival data and a diagnostic marker. *Biometrics*, 56(2):337–344, 2000.

[115] Ewout W Steyerberg et al. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer, 2008.

[116] Ewout W Steyerberg, Andrew J Vickers, Nancy R Cook, Thomas Gerds, Mithat Gonen, Nancy Obuchowski, Michael J Pencina, and Michael W Kattan. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.)*, 21(1):128, 2010.

[117] Olga V Demler, Nina P Paynter, and Nancy R Cook. Tests of calibration and goodness-of-fit in the survival setting. *Statistics in medicine*, 34(10):1659–1680, 2015.

[118] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[119] Jane Warwick, Lazlo Tabàr, Bedrich Vitak, and Stephen W Duffy. Time-dependent effects on survival in breast carcinoma: results of 20 years of follow-up from the swedish two-county study. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 100(7):1331–1336, 2004.

[120] David Collett. *Modelling survival data in medical research*. CRC press, 2015.

[121] Zhiliang Ying, Sin-Ho Jung, and Lee-Jen Wei. Survival analysis with median regression models. *Journal of the American Statistical Association*, 90(429):178–184, 1995.

[122] Song Yang. Censored median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association*, 94(445):137–145, 1999.

[123] A Senthilselvan. Penalized likelihood estimation of hazard and intensity functions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 49(2):170–174, 1987.

[124] Michal Abrahamowicz, Antonio Clampl, and James O Ramsay. Nonparametric density estimation for censored survival data: Regression-spline approach. *Canadian Journal of Statistics*, 20(2):171–185, 1992.

[125] Charles G Moertel, Thomas R Fleming, John S Macdonald, Daniel G Haller, John A Laurie, Phyllis J Goodman, James S Ungerleider, William A Emerson, Douglas C Tormey, John H Glick, et al. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine*, 322(6):352–358, 1990.

[126] Charles G Moertel, Thomas R Fleming, John S Macdonald, Daniel G Haller, John A Laurie, Catherine M Tangen, James S Ungerleider, William A Emerson, Douglass C Tormey, John H Glick, et al. Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage iii colon carcinoma: a final report. *Annals of internal medicine*, 122(5):321–326, 1995.

[127] Timothy H Holtz, Maya Sternberg, Steve Kammerer, Kayla F Laserson, Vija Riekstina, Evija Zarovska, Vija Skripconoka, Charles D Wells, and Vaira Leimane. Time to sputum culture conversion in multidrug-resistant tuberculosis: predictors and relationship to treatment outcome. *Annals of internal medicine*, 144(9):650–659, 2006.

[128] Sung-Min Kim, Junwoo Park, Sun Hee Kim, Su-Yeon Park, Jee Young Kim, Jung-Joon Sung, Kyung Seok Park, and Kwang-Woo Lee. Factors associated with the time to next attack in neuromyelitis optica: accelerated failure time models with random effects. *PLoS One*, 8(12), 2013.

[129] Eugenio Montini, Daniela Cesana, Manfred Schmidt, Francesca Sanvito, Cynthia C Bartholomae, Marco Ranzani, Fabrizio Benedicenti, Lucia Sergi Sergi, Alessandro Ambrosi, Maurilio Ponzoni, et al. The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of hsc gene therapy. *The Journal of clinical investigation*, 119(4):964–975, 2009.

[130] Nicholas F Marko, Robert J Weil, Jason L Schroeder, Frederick F Lang, Dima Suki, and Raymond E Sawaya. Extent of resection of glioblastoma revisited: personalized survival modeling facilitates more accurate survival prediction and supports a maximum-safe-resection approach to surgery. *Journal of clinical oncology*, 32(8):774, 2014.

[131] Lili Yu and Karl E Peace. Spline nonparametric quasi-likelihood regression within the frame of the accelerated failure time model. *Computational Statistics & Data Analysis*, 56(9):2675–2687, 2012.

[132] Trevor Hastie and Robert Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):757–779, 1993.

[133] Zongwu Cai and Yanqing Sun. Local linear estimation for time-dependent coefficients in cox's regression models. *Scandinavian Journal of Statistics*, 30(1):93–111, 2003.

[134] David M Zucker and Alan F Karr. Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *The Annals of Statistics*, pages 329–353, 1990.

[135] Michael J Crowther and Paul C Lambert. A general framework for parametric survival analysis. *Statistics in Medicine*, 33(30):5280–5297, 2014.

[136] James Robins and Anastasios A Tsiatis. Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika*, 79(2):311–319, 1992.

[137] Lynn Kuo and Bani Mallick. Bayesian semiparametric inference for the accelerated failure-time model. *Canadian Journal of Statistics*, 25(4):457–472, 1997.

[138] Patricia M Grambsch and Terry M Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526, 1994.

[139] Jianhua Z Huang and Linxu Liu. Polynomial spline estimation and inference of proportional hazards regression models with flexible relative risk form. *Biometrics*, 62(3):793–802, 2006.

[140] Loki Natarajan, Minya Pu, Barbara A Parker, Cynthia A Thomson, Bette J Caan, Shirley W Flatt, Lisa Madlensky, Richard A Hajek, Wael K Al-Delaimy, Nazmus Saquib, et al. Time-varying effects of prognostic factors associated with disease-free survival in breast cancer. *American journal of epidemiology*, 169(12):1463–1470, 2009.

[141] Hiroto Inaba, Harriet C Surprise, Stanley Pounds, Xueyuan Cao, Scott C Howard, Karen Ringwald-Smith, Jassada Buaboonnam, Gary Dahl, W Paul Bowman, Jeffrey W Taub, et al. Effect of body mass index on the outcome of children with acute myeloid leukemia. *Cancer*, 118(23):5989–5996, 2012.

[142] Jean-Pierre Quenot, Christine Binquet, Fady Kara, Olivier Martinet, Frederique Ganster, Jean-Christophe Navellou, Vincent Castelain, Damien Barraud, Joel Cousson, Guillaume Louis, et al. The epidemiology of septic shock in french intensive care units: the prospective multicenter cohort episs study. *Critical care*, 17(2):R65, 2013.

[143] Arnaud Pavon, Christine Binquet, Fady Kara, Olivier Martinet, Frederique Ganster, Jean-Christophe Navellou, Vincent Castelain, Damien Barraud, Joel Cousson, Guillaume Louis, et al. Profile of the risk of death after septic shock in the present era: an epidemiologic study. *Critical care medicine*, 41(11):2600–2609, 2013.

[144] PC Lambert, PW Dickman, CP Nelson, and P Royston. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Statistics in medicine*, 29(7-8):885–895, 2010.

[145] Kate E Oliver, William E Brady, Michael Birrer, David M Gershenson, Gini Fleming, Larry J Copeland, Krishnansu Tewari, Peter A Argenta, Robert S Mannel, Angeles Alvarez Secord, et al. An evaluation of progression free survival and overall survival of ovarian cancer patients with clear cell carcinoma versus serous carcinoma treated with platinum therapy: an nrg oncology/gynecologic oncology group experience. *Gynecologic oncology*, 147(2):243–249, 2017.

[146] Frank V Fossella, Russell DeVore, Ronald N Kerr, Jeffrey Crawford, Ronald R Natale, Frank Dunphy, Leonard Kalman, Vincent Miller, Jin Soo Lee, Melvin Moore, et al. Randomized phase iii trial of docetaxel versus vinorelbine or ifosfamide in patients with advanced non–small-cell lung cancer previously treated with platinum-containing chemotherapy regimens. *Journal of Clinical Oncology*, 18(12):2354–2362, 2000.

[147] Karynsa Cetin, David S Ettinger, Yong-jiang Hei, and Cynthia D O'Malley. Survival by histologic subtype in stage iv nonsmall cell lung cancer based on data from the surveillance, epidemiology and end results program. *Clinical epidemiology*, 3:139, 2011.

[148] Vincent T DeVita, Theodore S Lawrence, and Steven A Rosenberg. *Cancer of the Breast: From Cancer: Principles & Practice of Oncology*. Lippincott Williams & Wilkins, 2008.

[149] Jane C Weeks, E Francis Cook, Steven J O'Day, Lynn M Peterson, Neil Wenger, Douglas Reding, Frank E Harrell, Peter Kussin, Neil V Dawson, Alfred F Connors Jr, et al. Relationship between cancer patients' predictions of prognosis and their treatment preferences. *Jama*, 279(21):1709–1714, 1998.

[150] C Tibaldi, E Vasile, I Bernardini, C Orlandini, M Andreuccetti, and Alfredo Falcone. Baseline elevated leukocyte count in peripheral blood is associated with poor survival in patients with advanced non-small cell lung cancer: a prognostic model. *Journal of cancer research and clinical oncology*, 134(10):1143–1149, 2008.

[151] Khaled M Sarraf, Elizabeth Belcher, Evgeny Raevsky, Andrew G Nicholson, Peter Goldstraw, and Eric Lim. Neutrophil/lymphocyte ratio and its association with survival after complete resection in non–small cell lung cancer. *The Journal of thoracic and cardiovascular surgery*, 137(2):425–428, 2009.

[152] David P Carbone, Keyue Ding, Heinrich Roder, Julia Grigorieva, Joanna Roder, Ming-Sound Tsao, Lesley Seymour, and Frances A Shepherd. Prognostic and predictive role of the veristrat plasma test in patients with advanced non–small-cell lung cancer treated with erlotinib or placebo in the ncic clinical trials group br. 21 trial. *Journal of Thoracic Oncology*, 7(11):1653–1660, 2012.

[153] J Watine. Prognostic evaluation of primary non-small cell lung carcinoma patients using biological fluid variables. a systematic review. *Scandinavian journal of clinical and laboratory investigation*, 60(4):259–274, 2000.

[154] K Kato, Y Hitsuda, Y Kawasaki, T Igishi, K Yasuda, M Mikami, M Watanabe, M Miyata, T Sasaki, and E Shimizu. The value of serum c-reactive protein as a survival determinant in patients with advanced non-small-cell lung cancer. *Nihon Kokyuki Gakkai zasshi= the journal of the Japanese Respiratory Society*, 38(8):575–580, 2000.

[155] LM Forrest, DC McMillan, CS McArdle, WJ Angerson, and DJ Dunlop. Comparison of an inflammation-based prognostic score (gps) with performance status (ecog) in patients receiving platinum-based chemotherapy for inoperable non-small-cell lung cancer. *British journal of cancer*, 90(9):1704–1706, 2004.

[156] Daniel W Nixon, Steven B Heymsfield, Alice E Cohen, Michael H Kutner, Joseph Ansley, David H Lawson, and Daniel Rudman. Protein-calorie undernutrition in hospitalized cancer patients. *The American journal of medicine*, 68(5):683–690, 1980.

[157] SD Heys, KGM Park, PJ Garlick, and O Eremin. Nutrition and malignant disease: implications for surgical practice. *British journal of surgery*, 79(7):614–623, 1992.

[158] Leo Breiman et al. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231, 2001.

[159] Carine A Bellera, Gaëtan MacGrogan, Marc Debled, Christine Tunon de Lara, Véronique Brouste, and Simone Mathoulin-Pélissier. Variables with time-varying effects and the cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC medical research methodology*, 10(1):20, 2010.

[160] Michał Abrahamowicz and Antonio Ciampi. Information theoretic criteria in nonparametric density estimation: Bias and variance in the infinite dimensional case. *Computational statistics & data analysis*, 12(2):239–247, 1991.

[161] Torben Martinussen and Thomas H Scheike. A flexible additive multiplicative hazard model. *Biometrika*, 89(2):283–298, 2002.

[162] James M Robins and Dianne M Finkelstein. Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (ipcw) log-rank tests. *Biometrics*, 56(3):779–788, 2000.

[163] Cheryl L Faucett, Nathaniel Schenker, and Jeremy MG Taylor. Survival analysis using auxiliary variables via multiple imputation, with application to aids clinical trial data. *Biometrics*, 58(1):37–47, 2002.