MPL Estimation of a Proportional Hazards Mixture Cure Model with Partly-Interval Censoring

Annabel Webb

A thesis submitted to Macquarie University

for the degree of Master of Research

Department of Mathematics and Statistics

November 2020

MACQUARIE
University
SYDNEY·AUSTRALIA

The work presented here has not been submitted,
in whole or in part, for a higher degree to any
other university or institution.

Signed: [redacted]

Annabel Webb

Acknowledgements

This thesis could not have been completed without the support of everybody in the Department of Mathematics and Statistics. Most importantly, I would like to express my gratitude to my supervisor, A/Prof Jun Ma. I could not have conceived of completing this thesis without his guidance, input, feedback and expertise. I would also like to thank the Department and Faculty HDR staff, especially Richard Garner, Adam Sikora, and Eva Lu for their help and hard work. Acknowledgement must also go to the other academic staff in the Department who have offered me support and advice over the course of my undergraduate and postgraduate studies, especially Thomas Fung and Ayse Bilgen. I would also like to thank the MRes cohort in the Department, especially Jun Han, for going on this journey with me!

Thanks must also go to my friends, family, and loved ones. Thank you especially to Isaac Roberts and Gaby Barford, for keeping me sane during a very strange year, and always reminding me that I am more capable than I believe. Although the circumstances of this year have meant we have not been able to spend much time together, I would also like to acknowledge my extended family, especially my grandpa. Finally, I must acknowledge the overwhelming debt of gratitude I owe to my parents. I feel exceedingly lucky to have parents who have always valued, supported, and encouraged me and my education. This thesis certainly could not have been completed without their support during the toughest times of this extraordinary year.

Abstract

Proportional hazards mixture cure models are an important tool in survival analysis because they account for the presence of a sub-population who will never experience the event of interest. Much previous research in this area has been limited in scope to right censored data and has not offered a smooth estimate of the baseline hazard function. This thesis considers a maximum penalised likelihood (MPL) estimation of a proportional hazards mixture cure model for partly-interval censored survival data. The MPL method simultaneously estimates all model parameters, including a smooth M-spline approximation to the baseline hazard function. The non-negativity constraint on the baseline hazard function is guaranteed through the use of a multiplicative-iterative algorithm. Asymptotic properties are presented to allow for large sample inference on all parameters, including regression parameters and survival quantities. The results of two simulation studies are presented to demonstrate the method's performance, including a comparison to an existing method. A newly developed package for implementing the model in `R` is outlined and an example of its use is demonstrated with data from a melanoma study.

Contents

List of Tables

List of Figures

1 Introduction

Survival analysis is the area of statistics primarily concerned with the modelling of time-to-event data. This is a type of data where the main variable of interest is a measure of the time until some event of interest occurs. This length of time is often referred to as a survival time or a failure time, and the event may be referred to as a failure or a death. The most well-known method for modelling survival times as a function of a set of covariates is Cox's proportional hazards regression model (Cox 1972). This model is expressed in terms of a hazard function, which gives the risk of the event of interest occurring at a given time, conditional on the event of interest having not already occurred. Under the Cox model, the hazard function is a product of a non-parametric baseline hazard function, common to all individuals, and an exponential function involving all covariates and regression coefficients which modulates the risk of failure at a given time. This model is popular because, with the use of Cox's partial likelihood estimation method, it is possible to leave the unknown baseline hazard function arbitrary and simply estimate the value of the regression parameters, which greatly simplifies the estimation process (Cox 1975). Frequently, however, we may encounter situations where Cox's partial likelihood is insufficient for fitting a proportional hazards regression model, because an estimate of the baseline hazard function is required as part of parameter estimation.

One feature of time-to-event data that may preclude the use of Cox's partial likelihood is the presence of censored survival times in the dataset. Censoring occurs in survival data when the event time is not directly observed. The most common type of censoring is right censoring, where the event of interest has not occurred by the final time at which an individual has been observed. This occurs frequently in clinical studies, where subjects may be lost to the study or the study may end before all of the subjects have experienced the event of interest. If a set of survival data consists of event times and right censoring times, it is still possible to use Cox's partial likelihood to fit a proportional hazards model. However, this is not the case if there are individuals in the dataset who have been subject to left or interval censoring. Left censoring occurs in survival data when an individual has already experienced the event of interest prior to the first observation time or enrolment in the study. Interval censoring arises where the event of interest is known to have occurred between two observation times, but the event time was not observed exactly. We can refer to a dataset which contains any combination of event times and right, left, or interval censored survival times as partly-interval censored data. Estimating the regression parameters of a proportional hazards model for partly-interval censored data necessitates some estimation of the baseline hazard function. This therefore requires a more complex approach than Cox's partial likelihood. Many existing methods, such as Finkelstein (1986) and Huang (1996), have addressed this by treating the smooth baseline hazard as discrete or as a step function to simplify parameter estimation. Although this may be a straightforward approach, treating the baseline hazard function in this way limits the quality and interpretability of the baseline hazard function estimate produced.

Another complexity presented by some survival datasets is the possible presence of individuals who are not susceptible to ever experiencing the event of interest. This again prevents the straightforward estimation of a proportional hazards regression model using Cox's partial likelihood. The proportional hazards model assumes that all members of the population under consideration would eventually experience the event of interest if there was no restriction on follow-up time. However, there are a number of scenarios where it would be reasonable to assume that some members of the population may be cured of the underlying cause of the event, or otherwise have zero probability of ever experiencing it. This is an increasingly important factor to account for in survival analysis considering advancements in medical treatments and technologies over recent decades. Failing to account for this so-called cured fraction leads to overestimation of survival times. A natural extension to the proportional hazards model is the mixture cure model, which treats the population under consideration as a mixture of two subpopulations, where one is susceptible to the event and the other is not (Farewell 1982). Fitting this model becomes a two step process. The incidence, or probability of being susceptible to the event, is estimated using a method such as logistic regression. Then the latency, or time to the event, amongst the susceptible individuals is modelled using a method like proportional hazards regression. The form of the likelihood function for a mixture cure model using proportional hazards regression for the latency requires some estimation of the baseline hazard function in order to fit the model, even where the data is only subject to right censoring. This complexity is, perhaps, the reason that much of the previous work concerned with fitting a proportional hazards mixture cure model, such as Sy & Taylor (2000) and Peng & Dear (2000), has been limited in scope to right censored data. It is also notable that there has been virtually no consideration given to obtaining a smooth estimate of the baseline hazard function as part of fitting this model, even in more recent work that has incorporated partly-interval censored data (see Zhou et al. (2016), for example).

The use of a penalised likelihood approach could well provide a cohesive process for fitting a proportional hazards mixture cure model to partly-interval censored survival data and obtaining a smooth estimate of the baseline hazard function. A penalised likelihood makes use of the a priori knowledge that the baseline hazard function is smooth by introducing a roughness penalty term to the likelihood. It is therefore able to simultaneously produce regression parameter estimates for the latency and incidence models and a smooth estimate of the baseline hazard function. There is an existing body of research that has investigated the use of a penalised likelihood approach for estimating both the regression parameters and a smooth baseline hazard function for a proportional hazards model (Gray 1994, Joly et al. 1998, Cai & Betensky 2003, Ma et al. 2014). Key limitations of this aforementioned research have included, variously, restriction in scope to only right censored data, and an unsatisfactory treatment of the non-negative constraint on the baseline hazard function, leading to potential numerical issues. However, these limitations are cohesively addressed in the approach laid out in Ma et al. (2019). This research extends the multiplicative-iterative (MI) algorithm for constrained optimisation

developed in Chan & Ma (2012) to fit a proportional hazards regression model to partly-interval censored data. Despite the potential strengths of the penalised likelihood method in the context of proportional hazards regression model estimation, at present there has been a very limited amount of work done extending this approach to a mixture cure model. The only exception to this is Corbiere et al. (2009). Although this work showed promising results, it was limited in scope to right censored data, and also treated the non-negativity constraint on the baseline hazard function in a similarly ad-hoc manner to previous work in the field, again risking numerical issues in the estimates.

Given the aforementioned limitations of existing research, this thesis aims to construct and evaluate a maximum penalised likelihood (MPL) estimation of a proportional hazards mixture cure model that improves on Corbiere et al. (2009) by drawing from the methods presented in Ma et al. (2019). In doing so, this thesis will address the current lack of options for obtaining a smooth estimate of the baseline hazard function from a proportional hazards mixture cure model fitted to partly-interval censored data. By proposing a method for obtaining both regression parameter estimates and a smooth estimate of the baseline hazard function via a maximised penalised likelihood, this thesis will broaden the interpretability and utility of the proportional hazards mixture cure model for clinical and applied settings. Furthermore, this thesis will illustrate the utility of the algorithm presented by Ma et al. (2014) and Ma et al. (2019) for constrained optimisation in the context of a proportional hazards mixture cure model. In doing so, it will address the limitations of previous work which has treated the non-negativity constraint on the baseline hazard function in an ad-hoc manner, and thus avoid the potential for numerically unstable estimates. Additionally, in the course of evaluating the proposed estimation procedure, this project will develop an new option for fitting a proportional hazards mixture cure model in `R`. This is a significant contribution to the field, as existing options for practitioners who wish to fit this model in `R` are extremely limited.

This thesis is structured as follows. Chapter 2 presents a review of the existing literature. It will cover existing research concerned with fitting a proportional hazards regression model to partly-interval censored data, including obtaining a smooth estimate of the baseline hazard function, via a maximum penalised likelihood approach. Furthermore, it will outline previous approaches to fitting a proportional hazards mixture cure model, and discuss the ways in which the MPL method proposed here may be able to address the limitations of these approaches. It will also offer a brief overview of the currently limited options for fitting a proportional hazards mixture cure model in `R`. In Chapter 3, the model and estimation method under consideration in this thesis will be formally presented. This chapter will lay out the proportional hazards mixture cure model, discuss the parameterisation of the baseline hazard function and its approximation via basis functions, present the penalised likelihood function and detail the algorithm used to solve the constrained optimisation problem it presents. Additionally, this chapter will present asymptotic results that facilitate large sample inference using this model. Chapter 4 will present two simulation studies carried out to evaluate the performance of the proposed model. The

first of these is a comparative simulation study, contrasting the performance of the proposed method with an existing method for fitting the proportional hazards mixture cure model, using right censored data. The second is a simulation study using partly-interval censored survival data. Chapter 5 will provide an overview of the `R` package developed to fit the proposed model, and discuss an example of how the package could be used to analyse a real dataset. Finally, Chapter 6 will offer concluding remarks and comment on potential avenues for future research.

2 Literature Review

2.1 Penalised likelihood estimation of a proportional hazards model

Analysis of time-to-event data typically aims to model the time it will take for some event of interest to occur. This length of time is often referred to as a survival time. By far the most commonly used model in the analysis of survival times is the proportional hazards or Cox regression model (Cox 1972). A proportional hazards regression specifies the hazard function at time t as

$$h(t|\mathbf{x}_i) = h_0(t)\exp\{\mathbf{x}_i^T\boldsymbol{\gamma}\}$$

where $h_0(t)$ is the non-parametric baseline hazard function, $\mathbf{x}_i$ is a vector of covariates for the i-th individual and $\boldsymbol{\gamma}$ is a vector of regression coefficients. The popularity of the proportional hazards regression model stems largely from the fact that, by using Cox's partial likelihood, it is possible to estimate the regression coefficient vector $\boldsymbol{\gamma}$ without estimating the baseline hazard function (Cox 1975). This means that the effect of covariates on the hazard function can be estimated while the baseline hazard function remains unknown. However, Cox's partial likelihood is not suitable for fitting a proportional hazards regression model when survival times have been subject to left or interval censoring. This is because the baseline hazard function can no longer be avoided in the likelihood formulation. As such, alternative estimation methods are necessary.

One approach to estimating the baseline hazard function of a proportional hazards model for partly-interval censored data has been the use of a penalised likelihood (Ma et al. 2019). In this method, if $\boldsymbol{\eta}$ is the parameter vector of interest, then the aim is to estimate $\boldsymbol{\eta}$ by maximising the penalised log-likelihood function

$$\Phi(\boldsymbol{\eta}) = l(\boldsymbol{\eta}) - \lambda J(\boldsymbol{\eta})$$

where $l(\boldsymbol{\eta})$ is the log-likelihood function, $J(\boldsymbol{\eta})$ is a penalty function imposing smoothness, and λ is a non-negative smoothing parameter. The application an maximum penalised likelihood (MPL) approach in the context of the proportional hazards model is primarily motivated by the a priori knowledge that the non-parametric baseline hazard function is smooth and non-negative. Although a number of options for the roughness penalty function $J(\boldsymbol{\eta})$ exist, in practice the penalty function most commonly selected is $\int h_0''(v)^2 dv$, the L_2-norm of the second derivative of the baseline hazard function. The selection of L_2-norm is motivated by its appeal as an intuitive measure of the roughness of a function (Green & Silverman 1994).

An MPL estimation offers a number of benefits compared to alternative likelihood-based approaches to fitting a proportional hazards model to partly-interval censored survival data. Primarily, MPL estimation makes it possible to obtain a smooth estimate of the baseline hazard function. Alternative methods have generally not dealt with the baseline hazard function in such a satisfactory way. For instance, Finkelstein (1986) developed a method for testing

the regression coefficients of a proportional hazards model for partly-interval censored data, but considered the baseline hazard function as discrete. Huang (1996) likewise developed a likelihood based estimation of a proportional hazards model for interval censored data, but treated the cumulative baseline hazard function as a non-decreasing step function. Additionally, MPL estimation does not rely on bootstrapping for inference of the regression parameters, setting it apart from the method proposed by Pan (1999). Overall, a penalised likelihood method is an appealing approach to fitting a proportional hazards model to partly-interval censored data, particularly where one may be interested in estimating or carrying out inference on the baseline hazard function.

2.1.1 Approximation of the baseline hazard function

To obtain a smooth estimate of the baseline hazard function, it is necessary to address the fact that the function is dimensionally infinite (Ma et al. 2019). In MPL estimation, this is typically addressed by approximating the baseline hazard function using a set of basis functions and associated coefficients. As Ma et al. (2019) point out, basis functions can be selected from a wide range of choices, including indicator functions, Gaussian density functions and spline functions. In practice, cubic splines have proven a popular choice. For instance, Joly et al. (1998) made use of cubic M-splines (Ramsay 1988) in order to approximate the baseline hazard function, while Gray (1994) used cubic B-splines in his formulation of hypothesis tests for covariate effects and for the presence of proportional hazards. Conversely, Cai & Betensky (2003) used linear splines in their estimation of the baseline hazard function. However, linear splines offer less flexibility than cubic splines in this approximation process.

In addition to their greater flexibility, cubic splines, and cubic M-splines in particular, have other properties that make them well suited in the context of baseline hazard function approximation. Firstly, their formulation as a linear combination of piecewise polynomials means that guaranteeing a non-negative baseline hazard function approximation is simple. Each piecewise polynomial can only take positive or zero values (Ramsay 1988). This means that non-negativity can be ensured simply by restricting an associated set of linear coefficients to be non-negative. Secondly, each M-spline can be assigned an associated I-spline, its integral, and these I-splines will be monotonically increasing when associated with the same set of non-negative linear coefficients as the M-spline. This means that approximations of both the baseline hazard function and the cumulative baseline hazard function can be defined using the same vector of linear coefficients (Joly et al. 1998). The use of cubic M-splines for approximation therefore offers a flexible set of parameter estimates.

2.1.2 The non-negativity constraint and the multiplicative-iterative algorithm

A key property of the baseline hazard function is its strict non-negativity. As mentioned, if the baseline hazard function is to be approximated using a set of non-negative basis functions, it

is necessary to constrain the associated set of linear coefficients to be non-negative as well. In much of the aforementioned work, this element of baseline hazard function estimation has been either disregarded (Gray 1994) or not dealt with adequately. For instance, Joly et al. (1998) ensured a positive estimate for the baseline hazard function by squaring this set of coefficients. As pointed out by Ma et al. (2019), this has the potential to produce instability in the estimates and to create local maxima. Another approach was Cai & Betensky (2003)'s modelling of the log-baseline hazard, which leads to difficulties in obtaining a closed form for the cumulative baseline hazard and may also produce computational instability (Ma et al. 2014).

Some recent work has addressed this issue in a more rigorous fashion through the use of the multiplicative-iterative (MI) algorithm (Ma 2010, Chan & Ma 2012). This algorithm, which places a non-negative constraint on the estimation of the basis function coefficients as part of the estimation process, was developed in the context of image restoration using a penalised likelihood. Constrained optimisation arises in the field of statistical image restoration as pixel values of digital images can only be non-negative (Ma 2010, Chan & Ma 2012). The MI algorithm is developed from a decomposition of the gradient of the objective function into its positive and negative terms, allowing for the formulation of a step size that can guarantee a non-negative update (Ma 2010). The MI algorithm has a number of desirable properties for constrained optimisation in any context. Firstly, the algorithm updating scheme requires only the first derivative of the objective function, lessening the computational burden compared to some alternatives. Secondly, the ability to express the algorithm as a gradient based algorithm with an ascending direction allows for the easy incorporation of a line search step size (Ma 2010). This adds to the efficiency of parameter estimation as an increase in the objective function with each iteration is ensured.

This algorithm has been previously applied in the process of MPL estimation of a proportional hazards model in Ma et al. (2014) and Ma et al. (2019). Specifically, this work employed an alternating updating scheme, where the proportional hazard regression parameters were updated using the Newton-Raphson algorithm and the spline coefficients used in the baseline hazard approximation were updated using the MI algorithm. As a result, the non-negativity of the smooth baseline hazard function was guaranteed far more rigorously than in previous work.

2.1.3 Selection of the smoothing parameter λ

An additional element of maximum penalised likelihood estimation is the need to determine an appropriate value for the smoothing parameter λ. The value of λ controls the balance in the estimation process between the conflicting objectives of achieving a good fit to the data and obtaining a smooth estimate of the baseline hazard function. As Ma et al. (2019) observe, it is desirable for any MPL estimation process to include an automatic selection method for the smoothing parameter, as manual selection of λ is generally only appropriate for experienced

user. Inappropriate specification of λ may cause the method to be overly sensitive to the number of knots selected in the spline approximation to the baseline hazard function.

As pointed out by Wood (2011) and Krivobokova (2013), existing methods for selecting the optimal value of λ generally fall into one of two classes. The first of these classes is the set of methods concerned with selecting a value for λ that minimises some measure of model prediction error. These include methods that make use of a cross-validation or generalised cross-validation approach (Craven & Wahba 1979). This type of method is generally popular and has been made use of by some previous work in the area of MPL estimation of a proportional hazards model, such as Joly et al. (1998). Here, the smoothing parameter was selected by maximising an approximate cross-validation score, using a one-step Newton-Raphson expansion to reduce the computational burden of maximising the cross-validation directly. The second broad class of methods for selecting a value of the smoothing parameter tend to make use of a likelihood function for optimisation. The link between spline-smoothing and Bayesian estimation has been long established in the literature (Wahba 1978, Kimeldorf & Wahba 1970). As Krivobokova (2013) points out, this link can be exploited to allow for estimation of the smoothing parameter using a likelihood function. Within such a framework, the penalty term is thought of as a normal prior distribution and the smoothing parameter can be related to the variance of this distribution, meaning that likelihood methods can be used (Kauermann et al. 2009).

There has been fairly extensive work done considering the relative merits of these two groups of methods for selecting the smoothing parameter. Overall, existing result tend to suggest that likelihood-based methods are prone to under-smoothing in some cases, while generalised cross-validation methods tend to show greater variability in the smoothing parameter estimates obtained (Wahba 1985, Stein 1990, Kauermann 2005). However, results from Reiss & Ogden (2009) indicate that in finite samples generalised cross-validation is more likely to develop multiple minima and only weakly penalise overfitting compared to likelihood-based methods. A marginal likelihood based approach to selecting the smoothing parameter in an MPL estimation of a proportional hazards regression model has been used by Cai & Betensky (2003) and Ma et al. (2019). For deriving this method, it is common to employ a Laplace approximation so that an approximate marginal likelihood can be obtained (Wood 2011, Kauermann et al. 2009). As Ma et al. (2019) remark, such a method leads to quick convergence in the smoothing parameter estimate when an appropriate number of knots is selected for the spline approximation to the baseline hazard function.

2.2 Cure models in survival analysis

A feature of some time-to-event data that has been given consideration in survival analysis literature is the presence of long-term survivors. These are individuals who have not experienced the event of interest even after a lengthy follow-up time. Typical survival analysis methods, such as proportional hazards regression, assume that all members of the population under

consideration are susceptible to experiencing the event of interest, and that an event time would be observed for every individual given sufficient follow-up time. However, there are a number of scenarios where there is a reasonable possibility that some proportion of the population will never experience the event of interest. In these situations, disregarding the possibility that some individuals are not susceptible to the event leads to biased estimates of survival times. Methods that allow for the analysis of survival times while accounting for the presence of this so-called 'cured fraction' are therefore of interest, giving rise to cure models.

Cure models in survival analysis broadly fall into one of two categories. The most popular approach are mixture cure models, first proposed by Farewell (1982). These models consider the population of interest to be a mixture of two underlying sub-populations, one which is susceptible to the event of interest, and another which is not susceptible and will never experience the event. Fitting a model to these survival times becomes a two-step process of modelling the incidence, or the probability of an individual being in the non-cured fraction, and then the latency, or the time until the event of interest amongst the non-cured fraction. An alternative to mixture cure models is the promotion time model; see Tsodikov et al. (2003). Under such a model, the survival function is formulated such that it produces an estimate of the cured proportion if t is set to infinity, and produces a survival function for susceptible individuals otherwise (Zeng et al. 2006). Additionally, there has been a limited amount of work done on the development of a more general class of cure models, which includes the mixture cure model and the promotion time model as two special cases (Yin & Ibrahim 2005). However, mixture cure models remain the most popular approach to modelling survival data with a cure fraction, due to their intuitive and straightforward formulation and the flexibility offered by the two-component model structure (Banerjee & Carlin 2004)

2.2.1 Parametric and non-parametric models for the incidence

One of the key benefits of the mixture cure model is the flexibility it allows in the choice of models for both the incidence and the latency. Despite this theoretical flexibility, in practice the model chosen for the incidence part is most commonly a logistic regression model, following on from the early work of Farewell (1982). This popularity is likely because logistic regression is a straightforward and well-known method for modelling a binary outcome, such as whether or not an individual is a member of a cured sub-population. The use of logistic regression also allows the prediction of the cure probability to depend on a set of covariates, which is not a feature of all proposed methods for modelling the latency.

However, some non-parametric alternatives to logistic regression for modelling the incidence have been proposed in the literature. Some early work proposed estimators related to the Kaplan-Meier estimate of the survival function. For instance, Maller & Zhou (1992) suggested that the cure probability be set at one minus the minimum observed value of the Kaplan-Meier empirical distribution function. The work of Xu & Peng (2014) expanded on this to incorporate

kernel weighting into the cure probability estimation and allow for the effect of covariates to be considered. An alternative non-parametric model for the cure probability was proposed by Wang et al. (2012), who made use of a penalised expectation-maximisation algorithm within a smoothing spline analysis of variance framework to produce a fully non-parametric mixture cure model. This thesis will limit its consideration of incidence models to logistic regression.

2.2.2 Parametric and non-parametric models for the latency

The introduction of a cured fraction results in significant complexities in the fitting of a model for the latency. The key difficulty of fitting a latency model arises from the fact that, with the presence of a cured fraction, the baseline survival function can no longer be treated as a nuisance parameter in the process of estimating the regression parameters. As such, a variety of parametric and non-parametric models for the latency have been proposed in the literature.

An initial approach was to propose fully parametric models for the latency. In his introduction of the mixture cure model to the literature, Farewell (1982) made use of a Weibull distribution to model the time to the event of interest. Further research has variously proposed the use of the exponential, log-normal, Gompertz, and Burr XII distributions for modelling the latency (Ghittany et al. 1994, Gamel & Voggel 1997, Gordon 1990, Cantor & Shuster 1992, Shao & Zhou 2004). However, the use of fully parametric models for the latency may introduce undesirable limitations in the form of strong assumptions about the shape of the unknown baseline survival function. As a compromise between these limiting assumptions and the convenience of a parametric model, Peng et al. (1998) proposed the use of the more flexible generalised F distribution family for modelling the baseline survival function. However, the use of this family of distributions commonly gives rise to computational difficulties. These include difficulties evaluating the density and survival functions when shape and scale parameters become extreme and the unavailability of likelihood derivatives with respect to the shape parameters Peng et al. (1998).

Subsequent research has proposed models for the latency that aim to avoid parametric assumptions by making use of semi-parametric models for the baseline survival function, such as the proportional hazards model. As mentioned, the mixture cure model formulation prohibits the use of Cox's partial likelihood estimation due to the inability to treat the baseline hazard function as a nuisance parameter. As such, there have been a variety of proposed approaches to the task of regression parameter estimation for a proportional hazards mixture cure model. Kuk & Chen (1992) took a marginal likelihood approach in order to eliminate the unknown baseline survival function in the process of proportional hazards regression parameter estimation. In practice, Monte Carlo methods were used to approximate the maximum of the marginal likelihood and obtain regression parameter estimates. Fixed values of these regression parameter estimates were then used to maximise the full likelihood, allowing for the estimation of the baseline survival function. This method showed improved performance compared to a

Weibull model, but the use of Monte Carlo methods significantly increased the complexity and computational burden of parameter estimation. Additionally, as noted by Peng & Dear (2000), the total elimination of the baseline survival function from regression parameter estimation via the use of a marginal likelihood function leads to information loss in the Monte Carlo approximation. This approach was extended by Sy & Taylor (2000), who developed maximum likelihood estimation techniques for jointly estimating the logistic regression parameters for the incidence and proportional hazards regression parameters for the latency. This research dealt with the estimation of the nuisance baseline survival function by applying profile-likelihood and non-parametric likelihood techniques. Alternatively, Peng & Dear (2000) approached the estimation of the baseline survival function in this type of model using a discrete distribution with probability mass only at uncensored observations.

One key issue with the aforementioned research is a limitation in scope to only right censored data. At present, there is only a small amount of work concerning fitting a proportional hazards mixture cure model to partly-interval censored survival data. One example is Zhou et al. (2016), who made use of a multiple imputation approach to first obtain regression parameter estimates for the incidence and latency of a proportional hazards mixture cure model, and then used the Breslow estimator for the baseline survival function. The use of the more general class of transformation models for survival analysis in the context of a mixture cure model for interval censored data has also received some attention. Transformation models are a broad class of survival models which includes the proportional hazards model as a special case. Shen et al. (2019) used the proportional hazards model in their simulation study illustrating the use of a logistic transformation mixture cure model. Their method used an expectation-maximisation algorithm with re-parameterisation of the cure probability, and regarded the baseline hazard function as a step-function with non-negative jumps at particular times. A similar approach was taken by Chen et al. (2019), who likewise investigated the use of a transformation mixture cure model for partly-interval censored data using a non-decreasing, right-continuous step function for the baseline hazard function.

Despite the consideration of more complex censoring in more recent work on the proportional hazards mixture cure model, a common theme across much of the existing research is the unavailability of a smooth estimate of the baseline survival function. One of the few exceptions to this is the work done by Corbiere et al. (2009). In this research, the non-parametric estimator of the baseline survival function used was a smooth function maximising a penalised likelihood, which could be approximated via M-splines. This method thus provides a smooth estimate of the baseline hazard function. However, the work presented in Corbiere et al. (2009) still faces some key limitations. Firstly, it is restricted in scope to right-censored data. Additionally, the non-negativity constraint on the baseline hazard function is addressed in the ad-hoc manner of squaring the spline coefficient vector. As a result, this method suffers from identical drawbacks to the work discussed above in Section 2.1.2. Evidently, existing research that simultaneously offers a smooth estimate of the baseline hazard function, satisfactorily

addresses the non-negativity constraint, and can incorporate partly-interval censoring into the estimation of a proportional hazards mixture cure model is extremely limited.

2.2.3 The identifiability of the model & the zero-tail constraint

An issue commonly noted in the existing literature is the identifiability of mixture cure models, including those that adopt the proportional hazards model for the latency part. Many researchers have noted that the set of parameters that maximise the expected log-likelihood for such a model may not be unique (Amico & Van Keilegom 2018). Issues with the identifiability of this model arise where there exist right censoring times that are greater than the largest observed event time (Peng 2003). These are common in scenarios where a mixture cure model might be used, as the existence of these late right censoring times may be indicative of an underlying cured sub-population (Taylor 1995). However, these late right censoring times can result in estimates of the baseline survival function, such as those used in Sy & Taylor (2000) and Peng & Dear (2000), that do not approach 0 as $t \to \infty$. Consequently the baseline survival function for the non-cured population will be improper. If logistic regression is used for the incidence, there will be non-identifiability between the intercept parameter and the tail of this improper distribution (Taylor 1995, Peng 2003).

To address this issue of non-identifiability, many existing methods impose a zero-tail constraint on the baseline survival function. This constraint, which forces the survival function to zero at any t greater than the final event time, was first suggested by Taylor (1995) and has been included in much subsequent work including Sy & Taylor (2000) and Peng & Dear (2000). In essence, the constraint makes the strong assumption that all right censored observations after the final event time belong to the cured fraction. Research in the area of both parametric and non-parametric estimation of the cured proportion (Patilea & Van Keilegom 2017, Xu & Peng 2014) has noted that this assumption is crucial to ensure identifiability. Taylor (1995) argues that this assumption is reasonable, given that, as Farewell (1986) notes, it is generally only appropriate to fit a mixture cure model when follow-up time is sufficient to conclude that all subjects who are susceptible to the event of interest will have experienced it.

However, as pointed out by Peng (2003), this assumption may not be properly justified and may be inappropriate, especially when considering the practicalities of clinical research placing limits on follow-up time, or in the case of heavily censored samples (Corbiere et al. 2009). To address this, Peng (2003) proposed alternative methods for completing the tail of the baseline survival function. More specifically, this work proposed the use of a proper continuous distribution, such as an exponential or a Weibull distribution, to ensure that the estimate of the baseline survival function decreases to zero smoothly. Through the use of comparative simulation studies, Peng (2003) found that these alternatives could produce less biased estimates than the zero-tail constraint, improving estimation of covariate effects on the cured proportion and survival probabilities. However, these methods have not been widely

adopted; recent work such as Zhou et al. (2016) and Shen et al. (2019) has continued to make use of Taylor (1995)'s zero-tail constraint.

Evidently, in much of the previous literature on the estimation of a proportional hazards mixture cure model, the question of how to complete the tail of the unknown baseline survival function is a pertinent issue. However, results from the small amount of existing research on MPL estimation of this model indicate that, when this method is used, non-identifiability may not be as much of a concern. Specifically, Corbiere et al. (2009) modelled the baseline hazard function without any constraint on the right tail. They remark that poor choice of the smoothing parameter (see Section 2.1.3) may give an improper survival estimate that suggests an absence of a cured fraction in cases where there are very late uncensored event times. However, they also note that the presence of these very late event times may be an indication that a mixture cure model is not appropriate, either because follow-up was not sufficient or because there may in fact not be a cured fraction present in the population of interest. Overall, the results of Corbiere et al. (2009) suggest that a potential additional benefit to MPL estimation of a proportional hazards mixture cure model might be the ability to disregard concern about the identifiability of the model in practice.

2.3 Mixture cure model implementation in `R`

At present, options for fitting a proportional hazards mixture cure model in `R` are extremely limited. One available package is the `smcure` package (Cai et al. 2012*b*), which uses the expectation-maximisation algorithm outlined in Peng & Dear (2000) to estimate regression parameters for the latency and incidence of a mixture cure model. Both the proportional hazards regression model and the accelerated failure time model are available for the latency, and available link functions for the incidence are the logit link, probit link, and complementary log link. As such, the `smcure` package is an option for those wishing to fit a proportional hazards mixture cure model using logistic regression to model the incidence. However, the use of this package is limited to data where only right censoring is present. Furthermore, the standard errors for the regression parameters are obtained via bootstrapping, increasing the computational burden significantly. Finally, no smooth estimate of the baseline survival or hazard functions are available. Instead, a Breslow estimate of the baseline survival function is used to facilitate estimation of the regression parameters (Cai et al. 2012*a*).

At present, seemingly the only `R` implementation of a method to fit mixture cure models to partly-interval censored data is the `flexsurvcure` package (Amdahl 2020), which is a wrap-around to the `flexsurv` package (Jackson 2016). The `flexsurv` package allows users to fit a parametric regression model to survival data, by using one of several inbuilt parametric distributions or by specifying their own. Depending on the choice of parametric distribution, both proportional hazards regression and the accelerated failure time model are available. If one wishes to fit a proportional hazards model with only right censoring, the available

distributions are the exponential distribution and the Gompertz distribution; if one wishes to fit this model to partly-interval censored data, the Gompertz distribution must be selected. The package can then provide regression parameter estimates and inference on these parameters. `flexsurvcure` provides an extension to the methods available in `flexsurv` to fit a proportional hazards mixture cure model using logistic regression for the incidence. Available choices for the parametric baseline survival function are limited similarly in `flexsurvcure` as they are in `flexsurv`. The package provides regression parameters for the latency model and inference on these parameters, and an estimate of the size of the cured proportion, making use of the same set of covariates for both the latency and incidence models. Additionally, the package provides estimates of the parameters for the selected distribution, such as the shape and scale parameters for the Gompertz distribution. Overall, there is an obvious lack of options for easily fitting a proportional hazards mixture cure model within `R` that can account for partly-interval censored data, can provide a smooth estimate of the baseline hazard function, and is not limited by parametric assumptions.

3 Model Specification and Estimation Procedure

3.1 The proportional hazards mixture cure model for partly-interval censored data

The aim of the research presented here is to fit a proportional hazards mixture cure model to partly-interval censored survival data with a cured fraction. In order to incorporate event times that may be observed, right censored, left censored, or interval censored, event and censoring times are denoted as follows. For an individual i where $i = 1, ..., n$, Y_i is a random variable denoting the time until the event of interest, and $\mathbf{C}_i = (C_i^L, C_i^R)^T$ is a random vector denoting the respective end points of a random censoring interval. For this vector $\mathbf{C}_i$, we must have $C_i^R > C_i^L$ and $C_i^L \geq 0$, and it is possible to have $C_i^R = +\infty$. We assume that $\mathbf{C_i}$ and Y_i are independent and that they cannot be observed simultaneously. Given these conditions, our recorded survival time for each individual will consist of a random vector $\mathbf{T}_i = (T_i^L, T_i^R)^T$. If individual i is not subject to any type of censoring, we will record Y_i such that $T_i^L = T_i^R = Y_i$. Otherwise, we will record $\mathbf{C}_i$, giving us $T_i^L = C_i^L$ and $T_i^R = C_i^R$, and can say that $Y_i \in [C_i^L, C_i^R]$. We assume that the $\mathbf{T}_i$ are independent across the values of i, and denote the observed values of T_i^L and T_i^R respectively as t_i^L and t_i^R.

In addition to t_i^L and t_i^R, we also record two covariate vectors for each individual i, denoted as $\mathbf{x}_i$ and $\mathbf{z}_i$. $\mathbf{x}_i$ is the covariate vector used to model the latency of the mixture cure model, and $\mathbf{z}_i$ is the covariate vector used to model the incidence. The vector $\mathbf{x}_i$ is of length q while the vector $\mathbf{z}_i$ is of length p, and the two covariate vectors may be identical, may have no overlap, or one may be a subset of the other. Note that $\mathbf{x}_i$ forms the i-th row of the design matrix $\mathbf{X}$ and that $\mathbf{z}_i$ forms the i-th row of the design matrix $\mathbf{Z}$. Additionally, note that we may not necessarily record a variable denoting the censoring status of a given individual i. Nonetheless, we can easily compute a set of indicator variables that represent this information. This is because information about censoring is inherent in the recorded values of (t_i^L, t_i^R). Let δ_i be an indicator value for event times and δ_i^R, δ_i^L and δ_i^I be indicator values for right, left, and interval censoring respectively, such that $\delta_i = 1 - \delta_i^R - \delta_i^L - \delta_i^I$. In a case where $t_i^L = t_i^R$, the observed survival time is an event time and thus $\delta_i = 1$. If $t_i^R = +\infty$, then the survival time has been right censored, giving rise to $\delta_i^R = 1$. If $t_i^L = 0$, the survival time has been left censored, giving rise to $\delta_i^L = 1$. In any of these cases, we may denote the single observed time point as t_i in order to simplify notation. In any other case, the values of t_i^L and t_i^R indicate an interval censored survival time, giving rise to $\delta_i^I = 1$. We can say overall that the set of available information for each subject i is $(t_i^L, t_i^R, \mathbf{x}_i, \mathbf{z}_i, \delta_i, \delta_i^R, \delta_i^L, \delta_i^I)$.

Further to this observed information, let U_i be an unobserved random variable such that $U_i = 1$ if subject i is not in the cured fraction, and $U_i = 0$ if subject i is in the cured fraction. Note that although the value of u_i is unobserved, the values of t_i^L and t_i^R do offer information about the value of u_i for some i. Namely, if for a given subject i we have any case other

than $t_i^R = +\infty$ then we can conclude that $u_i = 1$. More clearly, we can say definitively that if the survival time of an individual has not been right censored, then they are not in the cured fraction. This follows from the fact that in the case of an observed, left censored, or interval censored survival time, the given individual is confirmed to have experienced the event of interest at some time either before or during the follow-up time. Conversely, if an individual has a right censored survival time, we have no information about the value of u_i. We cannot determine whether the event of interest did not occur by the final follow-up time because individual i is not susceptible to the event of interest, or if the event simply occurred after the end of observation.

From this, we can specify the survival function for the whole population, consisting of both the cured and non-cured fractions, as

$$S_{pop}(t|\mathbf{x}_i, \mathbf{z}_i) = \pi(\mathbf{z}_i)S(t|u_i = 1, \mathbf{x}_i) + (1 - \pi(\mathbf{z}_i))$$

where $S(t|u_i = 1, \mathbf{x}_i)$ is the survival function for the i-th subject, conditional on a set of covariates and the fact that subject i is not part of the cured fraction, and $\pi(\mathbf{z}_i)$ represents the probability that individual i is not part of the cured fraction given a set of covariates. We may refer to the function $S(t|u_i = 1, \mathbf{x}_i)$ as the conditional survival function, and, for ease of notation but without disregarding the conditionality on u_i, denote it as $S(t|\mathbf{x}_i)$. Further, we can define the probability density function for the whole population as

$$f_{pop}(t|\mathbf{x}_i, \mathbf{z}_i) = \pi(\mathbf{z}_i)f(t|u_i = 1, \mathbf{x}_i)$$

where $f(t|u_i = 1, \mathbf{x}_i)$ is the conditional probability density function for the i-th subject in the non-cured fraction, and can similarly be re-expressed as $f(t|\mathbf{x}_i)$ for ease of notation but without loss of conditionality. Finally, we can express the hazard function for the whole population as

$$\begin{aligned} h_{pop}(t|\mathbf{x}_i, \mathbf{z}_i) &= \frac{f_{pop}(t|\mathbf{x}_i, \mathbf{z}_i)}{S_{pop}(t|\mathbf{x}_i, \mathbf{z}_i)} \\ &= \frac{\pi(\mathbf{z}_i)f(t|\mathbf{x}_i)}{\pi(\mathbf{z}_i)S(t|\mathbf{x}_i) + (1 - \pi(\mathbf{z}_i))} \\ &= \frac{\pi(\mathbf{z}_i)S(t|\mathbf{x}_i)h(t|\mathbf{x}_i)}{\pi(\mathbf{z}_i)S(t|\mathbf{x}_i) + (1 - \pi(\mathbf{z}_i))} \end{aligned}$$

where $h(t|\mathbf{x}_i)$ is the conditional hazard function for the non-cured fraction.

Here, we wish to fit a proportional hazards model to the latency part of the mixture cure model. Under a proportional hazards model, the conditional survival function can be expressed as

$$S(t|\mathbf{x}_i) = S_0(t)^{\exp\{\mathbf{x}_i^T\boldsymbol{\gamma}\}}$$

where $S_0(t)$ is the baseline survival function and $\boldsymbol{\gamma}$ is a q-vector of proportional hazards regression parameters. More commonly, the proportional hazards model is expressed in terms of the hazard function. For the conditional hazard function, this expression is

$$h(t|u_i = 1, \mathbf{x}_i) = h(t|\mathbf{x}_i) = h_0(t)\exp\{\mathbf{x}_i^T\boldsymbol{\gamma}\}$$

where $h_0(t)$ is the baseline hazard function.

We also wish to model the incidence, or the probability that an individual is not in the cured fraction, using logistic regression. Specifically, we want to fit the the model

$$\pi(\mathbf{z}_i) = \frac{\exp\{\mathbf{z}_i^T\boldsymbol{\beta}\}}{1+\exp\{\mathbf{z}_i^T\boldsymbol{\beta}\}}$$

where $\boldsymbol{\beta}$ is a p-vector of regression parameters and $\mathbf{z}_i$ is as defined previously. As such, fitting this mixture cure model will require the estimation of the two regression parameter vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$.

3.2 Baseline hazard function approximation

In addition to estimating the proportional hazards regression parameter vector $\boldsymbol{\gamma}$ and the logistic regression parameter vector $\boldsymbol{\beta}$, this approach also estimates the baseline hazard $h_0(t)$. However, estimation of $h_0(t)$ without restriction using a finite number of survival times is ill-conditioned. Therefore, $h_0(t)$ must be approximated. The baseline hazard function can be approximated using some finite number m non-negative basis functions, such that

$$h_0(t) = \sum_{u=1}^{m} \theta_u \psi_u(t)$$

where θ_u is an element of an m-vector $\boldsymbol{\theta}$ and each $\psi_u(t)$ is a non-negative basis function.

3.2.1 Approximation via M-splines

In this model, the baseline hazard function will be approximated using cubic M-splines (see Section 2.1.1). M-splines can be characterised by their order, denoted by o. Each M-spline, $\psi_u^o(t)$, is a piecewise polynomial of degree $o-1$. These spline functions are completely defined by the sequence of knots selected. Here, we can define $\boldsymbol{\alpha}$ as a vector of knots with length n_κ. Further to this, we can define the vector $\boldsymbol{\alpha}^* = [\min(\boldsymbol{\alpha})\mathbf{1}_{o-1}^T, \boldsymbol{\alpha}^T, \max(\boldsymbol{\alpha})\mathbf{1}_{o-1}^T]^T$, where $\mathbf{1}_{o-1}$ is a vector of ones of length $o-1$. Then, when $t_{(1)} \leq t \leq t_{(n)}$, we can define an M-spline of order o as

$$\psi_u^o(t) = \frac{o}{o-1}\frac{\delta(\alpha_u^* \leq t < \alpha_{u+o}^*)}{\alpha_{a+o}^* - \alpha_u^*}\left[(t-\alpha_u^*)\psi_u^{o-1}(t) + (\alpha_{u+o}^* - t)\psi_{u+1}^{o-1}(t)\right]$$

when $o > 1$, and when $o = 1$, as

$$\psi_u^1(t) = \frac{\delta(\alpha_u^* \leq t < \alpha_{u+1}^*)}{\alpha_{a+1}^* - \alpha_u^*}$$

where α_u is the u^{th} element of the knots vector $\boldsymbol{\alpha}$ with $\alpha_u < \alpha_{u+1}$, and $\delta(\cdot)$ is an indicator function. Each $\psi_u^o(t)$ will be non-zero over the interval (α_u, α_{u+o}) and is zero outside of these o intervals.

The use of M-splines as the basis functions for the baseline hazard function approximation has a number of implications for the estimation of the coefficient vector $\boldsymbol{\theta}$. Firstly, and conveniently, their use means that the non-negative constraint on $h_0(t)$ can be satisfied simply by

ensuring that each element of $\boldsymbol{\theta}$ is non-negative. This is because each M-spline $\psi_u^1(t)$ is guaranteed to be non-negative. Secondly, the number of knots selected to define the M-spline basis functions, n_κ, will determine the dimension of the $\boldsymbol{\theta}$ vector, m. Although there is no explicit constraint on the size selected for n_κ, it generally varies with the sample size n. As a result, if we were to consider a case where $n \to \infty$, we could also have $m \to \infty$. This would give rise to significant issues in the estimation of the vector $\boldsymbol{\theta}$. In practice, it is sufficient to constrain m to be a finite number that is nonetheless allowed to vary with n. Further discussion of this issue is presented in Section 3.5.

3.3 The penalised likelihood function

The proposed method is to estimate the three parameter vectors, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ simultaneously by maximising a penalised likelihood function. For convenience, we can say that $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta})$. As noted above, any individual with a survival time that is directly observed, left censored, or interval censored is known to be in the non-cured fraction. This means that we can denote the contribution of these individuals to the overall penalised likelihood function directly in terms of the conditional density function, conditional survival function, or conditional baseline hazard function. Thus, under the assumption of independent censoring, we can express the contribution to the likelihood function of an individual i with $\delta_i = 1$ as

$$\pi(\mathbf{z}_i) f(t|\mathbf{x}_i) = \pi(\mathbf{z}_i) h(t|\mathbf{x}_i) S(t|\mathbf{x}_i),$$

the contribution of an individual i with $\delta_i^L = 1$ as

$$\pi(\mathbf{z}_i)(1 - S(t|\mathbf{x}_i)),$$

and the contribution of an individual with $\delta_i^I = 1$ as

$$\pi(\mathbf{z}_i)(S(t^L|\mathbf{x}_i) - S(t^R|\mathbf{x}_i)).$$

However, for an individual who is subject to right-censoring, it is unknown whether they are part of the cured fraction or not. As such, their contribution to the likelihood function can be expressed in terms of the population survival function. Their contribution can be denoted as

$$S_{pop}(t|\mathbf{x}_i, \mathbf{z}_i) = 1 - \pi(\mathbf{z}_i) + \pi(\mathbf{z}_i) S(t|\mathbf{x}_i).$$

Then we can say that the likelihood function for the whole sample is

$$L(\boldsymbol{\eta}) = \prod_{i=1}^{n} \big(\pi(\mathbf{z}_i) h(t_i|\mathbf{x}_i) S(t_i|\mathbf{x}_i)\big)^{\delta_i} \times \big(1 - \pi(\mathbf{z}_i) + \pi(\mathbf{z}_i) S(t_i|\mathbf{x}_i)\big)^{\delta_i^R}$$

$$\times \big(\pi(\mathbf{z}_i)(1 - S(t_i|\mathbf{x}_i))\big)^{\delta_i^L} \times \big(\pi(\mathbf{z}_i)(S(t_i^L|\mathbf{x}_i) - S(t_i^R|\mathbf{x}_i))\big)^{\delta_i^I}$$

and the log-likelihood is

$$l(\boldsymbol{\eta}) = \sum_{i=1}^{n} \big(\delta_i (\ln \pi(\mathbf{z}_i) + \ln h_0(t) + \mathbf{x}_i^T \boldsymbol{\gamma} + \ln S(t_i|\mathbf{x}))$$

$$+\delta_i^R \ln(1-\pi(\mathbf{z}_i)+\pi(\mathbf{z}_i)S(t_i|\mathbf{x}_i)) + \delta_i^L(\ln\pi(\mathbf{z}_i)+\ln(1-S(t_i|\mathbf{x}_i)))$$
$$+\delta_i^I(\ln\pi(\mathbf{z}_i)+\ln(S(t_i^L|\mathbf{x}_i)-S(t_i^R|\mathbf{x}_i)))\big)$$

This log-likelihood function is then penalised to obtain a smooth estimate for the baseline hazard function. The penalised likelihood is given by

$$\Phi(\boldsymbol{\eta}) = l(\boldsymbol{\eta}) - \lambda J(\boldsymbol{\eta})$$

where $J(\boldsymbol{\eta})$ is a roughness penalty function and $\lambda \geq 0$ is a smoothing parameter, the estimation of which is discussed in Section 3.4.1. The roughness penalty function used here is the L_2-norm of the second derivative of the baseline hazard function, namely $\int h_0''(v)^2 dv$. Given the baseline hazard function is $h_0(t) = \sum_{u=1}^m \theta_u \psi_u(t)$, we can conventiently express this roughness penalty as $J(\boldsymbol{\eta}) = \boldsymbol{\theta}^T \mathbf{R} \boldsymbol{\theta}$, where $\mathbf{R}$ is an $m \times m$ matrix with the (u,v)-th element given by $r_{u,v} = \int \psi_u''(t)\psi_v''(t)dt$. Maximising the penalised likelihood function to estimate the parameter vector $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta})$ requires derivation of the score vector and Hessian matrix components, which can be found in Appendix A. The maximisation of $\Phi(\boldsymbol{\eta})$ given the constraint that $\boldsymbol{\theta} \geq 0$ element-wise is outlined in the next section.

3.4 Estimation procedure

In order to fit our desired model, we wish to simultaneously produce MPL estimates of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$, denoted as $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}})$. Because the baseline hazard function must be non-negative, obtaining $\hat{\boldsymbol{\eta}}$ is a constrained optimisation problem that can be defined as

$$\hat{\boldsymbol{\eta}} = \max_{\boldsymbol{\theta} \geq 0} \Phi(\boldsymbol{\eta}) = \max_{\boldsymbol{\theta} \geq 0} \{l(\boldsymbol{\eta}) - J(\boldsymbol{\theta})\}$$

Given the constraint on the MPL estimate $\boldsymbol{\theta}$, we have the Karush-Kuhn-Tucker conditions

$$\frac{\partial \Phi(\boldsymbol{\eta})}{\partial \beta_t} = 0$$
$$\frac{\partial \Phi(\boldsymbol{\eta})}{\partial \gamma_j} = 0$$
$$\frac{\partial \Phi(\boldsymbol{\eta})}{\partial \theta_w} = 0 \text{ if } \theta_w > 0$$
$$\frac{\partial \Phi(\boldsymbol{\eta})}{\partial \theta_w} < 0 \text{ if } \theta_w = 0$$

This problem is solved iteratively using the Newton-MI algorithm (Ma et al. 2019). Let $\boldsymbol{\beta}^{(k)}$, $\boldsymbol{\gamma}^{(k)}$, and $\boldsymbol{\theta}^{(k)}$ be, respectively, the estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\theta}$ at iteration k. Also, for any function $a(x)$, define $a(x)^+$ as the function's positive components and $a(x)^-$ as the function's negative components, so that $a(x)^+ + a(x)^- = a(x)$. Iteration $k+1$ is obtained in a three step process as follows. Firstly, obtain $\boldsymbol{\beta}^{(k+1)}$ using a modified Newton algorithm:

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + \omega_1^{(k)} \left[-\frac{\partial^2 \Phi(\boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)}, \boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right]^{-1} \left[\frac{\partial \Phi(\boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)}, \boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\beta}} \right]$$

where $\omega_1 \in (0, 1]$ is the line search step size used to ensure that $\Phi(\boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}^{(k)}, \boldsymbol{\theta}^{(k)}) \geq \Phi(\boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)}, \boldsymbol{\theta}^{(k)})$. The value of the line search step size can be determined by using, for instance, Armijo's rule (Armijo 1966). Secondly, obtain $\boldsymbol{\gamma}^{(k+1)}$ using the same modified Newton algorithm:

$$\boldsymbol{\gamma}^{(k+1)} = \boldsymbol{\gamma}^{(k)} + \omega_2^{(k)} \left[-\frac{\partial^2 \Phi(\boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}^{(k)}, \boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} \right]^{-1} \left[\frac{\partial \Phi(\boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}^{(k)}, \boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\gamma}} \right]$$

where ω_2 is defined similarly to ω_1. Finally, obtain $\boldsymbol{\theta}^{(k+1)}$ via the multiplicative-iterative algorithm:

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \omega_3^{(k)} D^{(k)} \frac{\partial \Phi(\boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}^{(k+1)}, \boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}}$$

where ω_3 is defined similarly to ω_1 and ω_2 and $D^{(k)}$ is a diagonal $m \times m$ matrix with elements $\theta_w^{(k)} / d_w^{(k)}$ for $w = 1, ..., m$, and

$$d_w^{(k)} = \left[\frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \theta_w} \right]^- + \lambda \left[\frac{\partial J(\boldsymbol{\theta})}{\partial \theta_w} \right]^+ + \xi$$

Referring to Appendix A, we can see that in this case $d_w^{(k)}$ will be equal to

$$d_w^{(k)} = \delta_i \Psi_w(t_i) e^{\mathbf{x}_i^T \boldsymbol{\gamma}} + \delta_i^R \frac{\pi(\mathbf{z}_i) S(t_i|\mathbf{x}_i) \Psi_w(t_i)}{1 - \pi(\mathbf{z}_i) + \pi(\mathbf{z}_i) S(t_i|\mathbf{x}_i)} e^{\mathbf{x}_i^T \boldsymbol{\gamma}}$$

$$+ \delta_i^I \frac{S(t_i^L|\mathbf{x}_i) \Psi_w(t_i^L)}{S(t_i^L|\mathbf{x}_i) - S(t_i^R|\mathbf{x}_i)} e^{\mathbf{x}_i^T \boldsymbol{\gamma}} + \lambda \left[\frac{\partial J(\boldsymbol{\theta})}{\partial \theta_w} \right]^+ + \xi_w$$

Note that $\xi_w \geq 0$ is a small constant included simply to avoid the numerical issue of a zero denominator in the calculation of $D^{(k)}$ and does not have any impact on the final solution for $\boldsymbol{\theta}$.

3.4.1 Estimation of the smoothing parameter λ

A marginal likelihood method for the automatic selection of the smoothing parameter, previously outlined in Ma et al. (2019), can also be used for this model. In this method, the penalty function $J(\boldsymbol{\eta}) = \boldsymbol{\theta}^T \mathbf{R} \boldsymbol{\theta}$ is related to a normal prior distribution for the vector $\boldsymbol{\theta}$ parameterised by $\sigma_{\boldsymbol{\theta}}^2 = 1/2\lambda$, so that we have the distribution $N(0_{m \times 1}, \sigma_{\boldsymbol{\theta}}^2 \mathbf{R}^{-1})$. We can then obtain the log-posterior,

$$l_p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}) = -\frac{m}{2} \log \sigma_{\boldsymbol{\theta}}^2 + l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}) - \frac{1}{2\sigma_{\boldsymbol{\theta}}^2} \boldsymbol{\theta}^T \mathbf{R} \boldsymbol{\theta}$$

The marginal likelihood for $\sigma_{\boldsymbol{\theta}}^2$ may be difficult to obtain directly, and as such it is appropriate to approximate it using Laplace's method. Applying the Laplace approximation and substituting in the MPL estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$, we can obtain the approximated log-marginal likelihood for $\sigma_{\boldsymbol{\theta}}^2$,

$$l_m(\sigma_{\boldsymbol{\theta}}^2) \approx -\frac{m}{2} \log \sigma_{\boldsymbol{\theta}}^2 + l(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}) - \frac{1}{2\sigma_{\boldsymbol{\theta}}^2} \hat{\boldsymbol{\theta}}^T \mathbf{R} \hat{\boldsymbol{\theta}} - \frac{1}{2} \log \mid \hat{\mathbf{G}} + \mathbf{Q}(\sigma_{\boldsymbol{\theta}}^2) \mid$$

where $\hat{\mathbf{G}}$ is the negative Hessian matrix from $l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta})$ evaluated at the MPL estimates $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\theta}}$, and

$$\mathbf{Q}(\sigma^2_{\boldsymbol{\theta}}) = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{\sigma^2_{\boldsymbol{\theta}}}\mathbf{R} \end{bmatrix}$$

The solution for $\sigma^2_{\boldsymbol{\theta}}$ that maximises the approximation of $l_m(\sigma^2_{\boldsymbol{\theta}})$ is

$$\hat{\sigma}^2_{\boldsymbol{\theta}} = \frac{\hat{\boldsymbol{\theta}}^T \mathbf{R} \hat{\boldsymbol{\theta}}}{m - \nu}$$

where $\nu = \text{tr}\{(\hat{\mathbf{G}} + \mathbf{Q}(\hat{\sigma}^2_{\boldsymbol{\theta}}))^{-1}\mathbf{Q}(\hat{\sigma}^2_{\boldsymbol{\theta}})\}$, and can be considered equivalent to the model degrees of freedom. Given that the estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\theta}$ depend on $\sigma^2_{\boldsymbol{\theta}}$, the maximising solution for $\sigma^2_{\boldsymbol{\theta}}$ allows for the development of an iterative procedure with two steps. Firstly, with $\sigma^2_{\boldsymbol{\theta}}$ fixed, the corresponding MPL estimates for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ are obtained. Then, $\sigma^2_{\boldsymbol{\theta}}$ is updated using the newest values for $\hat{\sigma}^2_{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\theta}}$ on the right hand side of the maximising solution. These two steps are repeated until ν is stabilised. An example of the automatic selection of the smoothing parameter can be seen in Section 5.2.

3.5 Asymptotic properties and inference

Development of the asymptotic properties of the proposed model allows for large sample inference to be conducted without reliance on bootstrapping or other computationally intensive methods. Following from Ma et al. (2019), it is possible to demonstrate asymptotic consistency for the MPL estimates of both sets of regression parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and the baseline hazard function $h_0(t)$.

Let a and b be the minimum and maximum of all the observed survival times respectively, including interval censoring but excluding 0 and ∞. Then, let $C^r[a, b]$ be the set of functions that have r continuous derivatives over $[a, b]$. The parameter space for $\boldsymbol{\beta}$ can be given by $B = \{\boldsymbol{\beta} : |\beta_t| \leq C_1 < \infty, \forall t\}$. The parameter space for $\boldsymbol{\gamma}$ can be given by $G = \{\boldsymbol{\gamma} : |\gamma_j| \leq C_2 < \infty, \forall j\}$. The parameter space for $h_0(t)$ can be given by $A = \{h_0(t) : h_0 \in C^r[a, b], 0 \leq h_0(t)] \leq C_3 < \infty, \forall t \in [a, b]\}$. Therefore, the parameter space for $\boldsymbol{\tau} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, h_0(t))$ is $\boldsymbol{\Gamma} = \{\boldsymbol{\tau} : \boldsymbol{\beta} \in B, \boldsymbol{\gamma} \in G, h_0 \in A\}$. Before defining the MPL estimator of $\boldsymbol{\tau}$, it is necessary to account for the fact that this method estimates an approximation of $h_0(t)$. For convenience, the approximation can be denoted as $\tilde{h}_0(t) = \sum_{u=1}^{m} \theta_u \psi_u(t)$. The parameter space for $\tilde{h}_0(t)$ can be given by $A_n = \{\tilde{h}_0(t) : 0 \leq \tilde{h}_0(t) \leq C_4 < \infty, \forall t \in [a, b]\}$. Then the parameter space for $\boldsymbol{\tau}_n$ is $\boldsymbol{\Gamma}_n = \{\boldsymbol{\tau}_n : \boldsymbol{\beta} \in B, \boldsymbol{\gamma} \in G, \tilde{h_0} \in A_n\}$. The MPL estimator of $\boldsymbol{\tau}_n$ is then $\hat{\boldsymbol{\tau}}_n = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{h}_0(t))$. Theorem 1 demonstrates asymptotic consistency for $\hat{\boldsymbol{\tau}}_n$ when the number of basis functions $m \to \infty$ but $m/n \to 0$ when $n \to \infty$, and the scaled smoothing value $\mu_n = \lambda/n \to 0$ when $n \to \infty$.

Theorem 1. *Assume that $h_0(t)$ is bounded and has some number $r \geq 1$ derivatives over the interval $[a, b]$. Assume that $m = n^v$, where $0 < v < 1$, and $\mu_n \to 0$ as $n \to \infty$. Then, when $n \to \infty$,*

1. $||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|| \to 0$ *almost surely, and*

2. $||\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0|| \to 0$ *almost surely, and*

3. $\sup_{t\in[a,b]} |\hat{h}_0(t) - h_{00}(t)| \to 0$ *almost surely.*

Theorem 1 requires the following regularity conditions:

1. The matrices $\mathbf{X}$ and $\mathbf{Z}$ are bounded, and both $E(\mathbf{X}\mathbf{X}^T)$ and $E(\mathbf{Z}\mathbf{Z}^T)$ are non-singular.

2. The penalty function $J(\boldsymbol{\eta})$ is bounded over $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_n$.

3. For function $\tilde{h}_0(t)$, there is a constant C_5 independent of n that is the upper bound of all $\theta_u \geq 0$. Additionally, the basis functions $\psi_u(t)$, where $u = 1, ..., m$, are bounded for $t \in [a, b]$.

4. The knots and basis functions are selected such that for any $h_0(t) \in A$ there is a $\tilde{h}_0(t) \in A_n$ which satisfies $\max_t |\tilde{h}_0(t) - h_0(t)| \to 0$ when $n \to \infty$.

Define the distance measure $\rho(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ as

$$\rho(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \{\|\boldsymbol{\tau}_1 - \boldsymbol{\tau}_1\|^2\}^{1/2} = \left\{ \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|_2^2 + \|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2\|_2^2 + \sup_{t\in[a,b]} |h_{01}(t) - h_{02}(t)|^2 \right\}^{1/2}$$

Under the above regularity conditions, Theorem 1 can be demonstrated by showing that $\rho(\boldsymbol{\tau}_0, \hat{\boldsymbol{\tau}}_n) \to 0$ almost surely, where $\boldsymbol{\tau}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, h_{00}(t))$ is the true parameter value. The required result can be obtained by applying Theorem 1 from Ma et al. (2019).

Additionally, it is desirable to jointly develop asymptotic normality results for all three parameters, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$. This allows for inference to be made not only on regression parameters but also on other quantities, such as survival probabilities. In order to develop these results, it is necessary to address the issue that, if we have $n \to \infty$, it is possible for $m \to \infty$ as well. Following Ma et al. (2019) and Yu & Ruppert (2002), this is addressed here by restricting m to be a finite number which is nonetheless allowed to vary with n. Furthermore, the development of asymptotic normality results for the parameters must take into account the possibility of encountering active constraints in the estimation of $\boldsymbol{\theta} \geq 0$, where θ_u may be equal to 0 for some u. This is particularly likely to occur when the number of knots is larger than strictly necessary, as the penalty function will push these unnecessary parameters to zero (Ma et al. 2019). It is important to take the potential presence of active constraints in the asymptotic covariance matrix into consideration as not doing so may produce undesirable results, such as negative variances.

Recall that we have defined the parameter vector $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta})$, which has a finite length of some number $p + q + m$, and that we can express the penalised likelihood function in terms of $\boldsymbol{\eta}$ such that

$$\Phi(\boldsymbol{\eta}) = l(\boldsymbol{\eta}) - \lambda J(\boldsymbol{\eta})$$

We denote the MPL estimate of $\boldsymbol{\eta}$, which is obtained by maximising $\Phi(\boldsymbol{\eta})$ with the constraint $\boldsymbol{\theta} \geq 0$, as $\hat{\boldsymbol{\eta}}$. Let the true value of $\boldsymbol{\eta}$ be represented by $\boldsymbol{\eta}_0$. Assume that we have a case where the first r elements of $\boldsymbol{\theta}$ are subject to active constraints in the MPL solution, and define

$$\mathbf{U} = [\mathbf{0}_{(m-r+p+q)\times r}, \mathbf{I}_{(m-r+p+q)\times(m-r+p+q)}]^T$$

where $\mathbf{0}$ is a matrix of zeros, $\mathbf{I}$ is an identity matrix, and $\mathbf{U}^T\mathbf{U} = \mathbf{I}_{(m-r+p+q)\times(m-r+p+q)}$ is satisfied.

Theorem 2. *Assume that $\mu_n = o(n^{1/2})$ and that we have r active constraints in the MPL estimate of $\boldsymbol{\theta}$. Define matrix $\mathbf{U}$ as above. Let*

$$\mathbf{F}(\boldsymbol{\eta}) = -E_{\boldsymbol{\eta}_0}\left[\lim_{n\to\infty} n^{-1}\frac{\partial^2 l(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}\right]$$

Under these conditions, when $n \to \infty$, $\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ converges in distribution to $\mathcal{N}(\mathbf{0}, \tilde{\mathbf{F}}(\boldsymbol{\eta}_0)^{-1})$, where $\tilde{\mathbf{F}}(\boldsymbol{\eta}_0)^{-1} = \mathbf{U}(\mathbf{U}^T\mathbf{F}(\boldsymbol{\eta})\mathbf{U})^{-1}\mathbf{U}^T$.

Theorem 2 can be shown under the following regularity conditions:

1. The distributions of $\mathbf{x}_i$ and $\mathbf{z}_i$ are independent of $\boldsymbol{\eta}$.

2. The limit $\lim_{n\to\infty}[n^{-1}l(\boldsymbol{\eta})]$ exists and has a unique maximum at $\boldsymbol{\eta}_0 \in \Omega$, where Ω is the parameter space for $\boldsymbol{\eta}$ and is a compact subspace of $\mathbf{R}^{p+q+m}$. That is to say, if the sample size is infinity, the true parameters can be obtained exactly from maximising the likelihood.

3. $l(\boldsymbol{\eta})$ has a finite upper bound and is twice continuously differentiable in a neighbourhood of $\boldsymbol{\eta}_0$, and the matrices
$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\frac{\partial l_i(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\frac{\partial l_i(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}^T}$$
and
$$\lim_{n\to\infty}\left[-n^{-1}\frac{\partial^2(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}\right]$$
exist.

4. The penalty function $J(\boldsymbol{\eta})$ is twice continuously differentiable on Ω, and these derivatives are bounded.

5. The matrix $\mathbf{U}^T\mathbf{F}(\boldsymbol{\eta})\mathbf{U}$ is invertible in a neighbourhood of $\boldsymbol{\eta}_0$.

If these conditions hold, it is a simple matter to demonstrate that $\hat{\boldsymbol{\eta}} \to \boldsymbol{\eta}_0$. Let $\bar{l}(\boldsymbol{\eta}) = \lim_{n\to\infty}[n^{-1}l(\boldsymbol{\eta})]$ exist with a unique maximum at $\boldsymbol{\eta}_0 \in \Omega$, where Ω is the parameter space for $\boldsymbol{\eta}$. Under the strong law of large numbers, we have $n^{-1}l(\boldsymbol{\eta}) \to \bar{l}(\boldsymbol{\eta})$ almost surely and uniformly for $\boldsymbol{\eta} \in \Omega$. Additionally, we have $\mu_n \to 0$ as $n \to \infty$. This is sufficient to show that $\hat{\boldsymbol{\eta}} \to \boldsymbol{\eta}_0$.

The asymptotic normality result can be proven by following Theorem 2 from Ma et al. (2019). Given that, according to the KKT conditions outlined above in Section 3.4, we have a constrained MPL estimate $\hat{\boldsymbol{\eta}}$ that satisfies

$$\mathbf{U}^T \frac{\partial \Phi(\hat{\boldsymbol{\eta}})}{\partial \boldsymbol{\eta}} = 0$$

it is possible to show that

$$\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) = -\mathbf{U}\left(\mathbf{U}^T \frac{1}{n} \frac{\partial^2 \Phi(\tilde{\boldsymbol{\eta}})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} \mathbf{U}\right)^{-1} \mathbf{U}^T \left(\frac{1}{\sqrt{n}} \frac{\partial l(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}} + o(1)\right)$$

where $\tilde{\boldsymbol{\eta}}$ is a vector between $\hat{\boldsymbol{\eta}}$ and $\boldsymbol{\eta}_0$. Here, when $n \to \infty$ and $\mu_n \to 0$, $n^{-1}\partial^2 \Phi(\tilde{\boldsymbol{\eta}})/\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T$ converges almost surely to $F(\boldsymbol{\eta}_0)$ under the law of large numbers. If we then apply the central limit theorem to $n^{-1/2}\partial l(\boldsymbol{\eta}_0)/\partial \boldsymbol{\eta}$, the asymptotic normality result is demonstrated.

In order to implement the result of Theorem 2, it is necessary to define a method for identifying active constraints when they arise in the MPL estimation of $\boldsymbol{\theta}$. The method used here follows that proposed by Ma et al. (2019). Active constraints can be identified by inspecting both the value of $\hat{\theta}_u$ and the corresponding gradient for each u. After the Newtown-MI algorithm has reached convergence, some $\hat{\theta}_u$ may be exactly zero with negative gradients, and thus are clearly subject to an active constraint. Furthermore, there may be some $\hat{\theta}_u$ that are very close to, but not exactly, zero. For these $\hat{\theta}_u$, a corresponding negative gradient value is indicative that they are also subject to an active constraint. In practice, active constraints are defined where, for a given u, $\hat{\theta}_u < 10^{-2}$ and the corresponding gradient is $< -10^{-2}$. After the indices associated with active constraints are identified, obtaining the matrix $\tilde{\mathbf{F}}(\boldsymbol{\eta}_0)^{-1}$ is a very straightforward computation, as Ma et al. (2019) point out. The matrix $\mathbf{U}^T\mathbf{F}(\boldsymbol{\eta})\mathbf{U}$ is obtained by removing the rows and columns of $\mathbf{F}(\boldsymbol{\eta})$ associated with the active constraints. The result is then inverted, and then padded with zeros in the deleted rows and columns to obtain $\tilde{\mathbf{F}}(\boldsymbol{\eta}_0)^{-1}$.

To make use of these asymptotic results for inference on finite samples, it is necessary to approximate the distribution for $\hat{\boldsymbol{\eta}}$ when n is large. Doing so also incorporates non-zero values for the smoothing parameter λ into the inference on the parameter estimates. The necessary results are presented below in Corollary 1.

Corollary 1. *Assume that the smoothing parameter $\lambda \ll n$. Define*

$$\mathbf{A}(\hat{\boldsymbol{\eta}})^{-1} = \mathbf{U}\left(\mathbf{U}^T\left(\frac{\partial^2 l(\hat{\boldsymbol{\eta}})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} + \lambda \frac{\partial^2 J(\hat{\boldsymbol{\eta}})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T}\right)\mathbf{U}\right)^{-1} \mathbf{U}^T$$

Then, when n is large, the distribution for the MPL estimate $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0$ can be approximated by a multivariate normal distribution having mean zero and covariance matrix

$$\hat{var}(\hat{\boldsymbol{\eta}}) = \mathbf{A}(\hat{\boldsymbol{\eta}})^{-1} \frac{\partial^2 l(\hat{\boldsymbol{\eta}})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} \mathbf{A}(\hat{\boldsymbol{\eta}})^{-1}$$

These results allow for inferences to be made not only on both sets of regression parameters but also on quantities associated with the baseline hazard function.

4 Simulation Studies

Two simulation studies were carried out to evaluate the performance of the proposed model. The key focus of these simulation studies was the investigation of

- the performance of the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$,
- the performance of the estimate for $h_0(t)$ and related survival quantities, and
- the performance of the asymptotic variance estimator for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $h_0(t)$.

The first simulation study considered right censored survival data with a cured fraction. This was a comparative study between the proposed MPL method and an existing method for fitting a proportional hazards mixture cure model to right censored data. Results and discussion regarding this study are presented in Section 4.1. The second simulation study considered partly-interval censored cure survival data. This was not a comparative study as there are presently limited alternatives for fitting a proportional hazards mixture cure model to partly-interval censored data in `R`. Results and discussion regarding this study can be found in Section 4.2.

4.1 Comparative simulation study using right censored data

The aim of the first simulation study was to compare the performance of the proposed method for fitting a proportional hazards mixture cure model with an existing alternative. Due to the limitations of available alternatives, this simulation study considered only right censored survival data with a cured fraction. This allowed for a comparison to be made between the proposed method and the `smcure` package, discussed previously in Section 2.3. The performance of the EM algorithm used in the `smcure` package was compared with the proposed method on the basis of both the regression parameter estimates (Section 4.1.1) and the estimates of the cumulative baseline hazard function (Section 4.1.2).

T ensure that the model was evaluated thoroughly, the simulation included a range of different sample sizes, cured fraction sizes, and event probabilities in the non-cured fraction. Table 1 summarises the different specifications considered in this first simulation. For each scenario, 500 samples were generated. The simulation involved two different cured fraction sizes and used a single binomial covariate, z_1, in the logistic regression for the cure probability. Following Corbiere et al. (2009), the size of the cured fraction in a given scenario was specified by setting the value of the intercept in the logistic regression, β_0, as this parameter controls the cure rate amongst individuals with $z_1 = 0$. Note that a value of $\beta_0 = 1$ corresponds to a non-cured proportion of approximately 0.75 in this group and a value of $\beta_0 = 0$ corresponds to a non-cured proportion of approximately 0.5.

Prior to generating observed survival times, an indicator value u_i was obtained for each

Table 1: Specifications for Simulation Study 1

Simulation Parameters	
$\boldsymbol{\gamma}$ vector	$[0.5]$
$\mathbf{X}$ vector	$\mathbf{X} = [x_1]$
Y distribution	Weibull
Baseline hazard	$h_0(y) = 3y^2$
Cured Fraction Scenarios	
$\boldsymbol{\beta}$ vector	$[1, 1]^T$ giving $\pi(\mathbf{z}) \simeq 0.75$ i.e. 75% of observations non-cured $[0, 1]^T$ giving $\pi(\mathbf{z}) \simeq 0.5$ i.e. 50% of observations non-cured
$\mathbf{Z}$ vector	$\mathbf{Z} = [1, z_1]$
Simulation Scenarios	
Sample size	$n = 100, 500, 2000$
Censoring distribution	$\text{Exp}(\lambda_c = 1.25)$ giving 50% censoring rate in non-cured fraction $\text{Exp}(\lambda_c = 4.2)$ giving 20% censoring rate in non-cured fraction

Note that the size of the cured fraction and the censoring rate in the non-cured fraction contribute to the rate of right censoring in the sample as a whole.

individual i so that

$$u_i = \begin{cases} 0 & \text{if } U_i^C > \pi(\mathbf{z}_i) \\ 1 & \text{if } U_i^C \leq \pi(\mathbf{z}_i) \end{cases}$$

where U_i^C denotes a standard uniform random variable and $\pi(\mathbf{z}_i)$ is

$$\pi(\mathbf{z}_i) = \frac{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{z}_i^T \boldsymbol{\beta})}$$

A value of $u_i = 0$ indicates that an individual is in the cured fraction, and $u_i = 1$ indicates that an individual is susceptible to the event of interest. Observed times for the cured fraction, all of which constituted right censoring times, were drawn from an exponential distribution.

For individuals in the non-cured fraction, where $u_i = 1$, event times were drawn from a Weibull distribution. Observed survival times T_i for individuals in the non-cured fraction, including both event and right censoring times, were obtained by

$$T_i = Y_i^{\delta(Y_i < C_i)} C_i^{\delta(C_i \geq Y_i)}$$

where Y_i is an event time drawn from a Weibull distribution, C_i is a censoring time drawn from an exponential distribution, and $\delta(\cdot)$ denotes an indicator function. A single binomial covariate, x_1, was used in the proportional hazards regression; this covariate was independent from that used in the logistic regression.

In this simulation study, two different values for the censoring rate in the non-cured fraction were considered, as shown in Table 1. The censoring rate in the non-cured fraction was

controlled by adjusting the parameter λ_c of the exponential function from which C_i was drawn. The two values of λ_c considered in this simulation study were $\lambda_c = 1.25$ and $\lambda_c = 4.2$. These corresponded to censoring rates in the non-cured fraction of approximately 50% and 20% respectively. However, these specifications did not correspond to the rate of censoring in the sample as a whole. As discussed, only right censoring times are observed for individuals in the cured fraction. This means that two parameters effect the extent of censoring in the sample as a whole:

- the event probability amongst the non-cured individuals π^E, and
- the size of the cured fraction $1 - \pi(\mathbf{z})$, controlled by the value of β_0.

This must be taken into consideration when evaluating the results of this simulation study. In the results that follow, the value reported for π^R is the proportion of right censoring in the sample as a whole, accounting for both the cured fraction and censoring in the non-cured fraction.

For each generated sample, cubic M-splines with some number n_κ knots were used to approximate the baseline hazard function. Define a and b similarly to their definition in Section 3.5, so that they are respectively the minimum and maximum of all observed survival times, including interval censoring but excluding 0 and ∞. The first and last knot, also referred to as the external knots, were placed at a and b respectively. The remaining knots, or the internal knots, were then placed at equal quantiles across the interval between $t_{0.075}$ and $t_{0.9}$, which correspond respectively to the 7.5^{th} percentile and the 90^{th} percentile between a and b. For each simulation scenario, it was necessary to determine an appropriate number of knots to use in the construction of the cubic M-splines. As Corbiere et al. (2009) remark, there is little practical benefit to increasing the number of knots beyond what is necessary to produce satisfactory parameter estimates. Previous work that has carried out simulation studies with comparable methods to the one presented here, such as Ma et al. (2019), has discussed the need for the number of knots simply to be allowed to vary with the sample size n, using, for instance, a rough guideline of the cubic root of the sample size rather than any strict optimisation process.

In this simulation study, the number of knots was likewise selected manually, rather than through any optimisation process. The procedure set out in work such as Ma et al. (2019) offered some guidance in the initial knot selection process. However, the process of determining the number of knots needed for this particular simulation study was complicated by the presence of the cured fraction, which was not a feature of the work done in Ma et al. (2019). Namely, changes in the size of the cured fraction resulted in varying convergence times when the sample size, number of knots, and all other simulation parameters were kept constant. In order to achieve reasonable convergence times for all simulation scenarios, it was therefore necessary to reduce the number of knots selected as the size of the cured fraction increased. The minimum number of knots used for any simulation scenario was $n_\kappa = 3$, corresponding to the two external knots and one internal quantile knot located at the 50^{th} percentile of $[t_{0.075}, t_{0.9}]$. The number of

knots selected for each simulation scenario can be found in Appendix B1. For the purposes of this simulation study, the value of the smoothing parameter λ was set manually to zero. That is to say, the method for automatic selection of the smoothing parameter laid out in Section 3.4.1 was not evaluated as part of this simulation study.

4.1.1 Regression parameter estimation results

The biases in both the MPL and EM estimates of β_0, β_1 and γ_1 are presented in Table 2. It is clear that both methods produce estimates of β_0 that have significant bias. Interestingly, in scenarios with the lowest proportion of right censoring, the MPL method produces estimates with systematic positive bias, while the EM method produces estimates with systematic negative bias. This is equivalent to the MPL method systematically underestimating the size of the cured fraction when right censoring is low, and the EM method systematically overestimating it. When the proportion of right censoring increases, both methods produce estimates with significant negative bias, meaning they are both systematically overestimating the size of the cured fraction. However, the magnitude of the systematic negative bias in $\hat{\beta}_0$ produced by the MPL method is consistently smaller than that produced by the EM method.

In the case of the MPL method, it is possible that the bias in $\hat{\beta}_0$ arises from issues with identifiability, as this is a recurrent issue in mixture cure model estimation (see Section 2.2.3). Corbiere et al. (2009) remark that issues with identifiability may not be as prevalent when using an MPL method to fit a mixture cure model, but that they may arise where the value of the smoothing parameter λ is inappropriate. It is therefore possible that there would have been reduced biased in the MPL estimates of β_0 if the iterative process for selecting the smoothing parameter laid out in Section 3.4.1 had been used. However, identifiability is assured for the `smcure` method, as a zero-tail constraint is imposed as part of the estimation process (Cai et al. 2012*a*). The bias in the EM estimates is therefore certainly not a result of issues with identifiability. Given this, it is possible that the bias in both the MPL and EM estimates of β_0 has resulted from of a lack of separation between the cured and non-cured populations in the data. The relatively large concentration of right censored observations in the sample as a whole across all simulation scenarios may have created difficulties distinguishing between cured and non-cured individuals. Evidently, estimation of the logistic regression intercept parameter is difficult and needs further investigation. For this reason, the rest of this section will consider only results pertaining to the estimates of β_1 and γ_1, including in the discussion of standard error estimation.

Results for the bias in the estimates of β_1 and γ_1 produced by both methods (Table 2) show that the performance of the two methods is generally comparable for these parameters, with both methods producing reasonably unbiased estimates. Table 3 presents the results of the standard error estimation for both methods. Comparison of the standard error estimation is especially useful as it is provides insight into the relative merits and weaknesses of estimating

Table 2: Bias in estimates of β_0, β_1 and γ_1

n		100	100	500	500	2000	2000
$\pi(\mathbf{z})$		0.5	0.75	0.5	0.75	0.5	0.75
π^R		0.59	0.41	0.59	0.41	0.59	0.41
$\hat{\beta}_0$	MPL	0.2531	0.3648	0.2970	0.3267	0.2961	0.3078
	EM	-0.2997	-0.1801	-0.2192	-0.2037	-0.2209	-0.2269
$\hat{\beta}_1$	MPL	0.0947	0.1027	0.0412	0.0633	0.0282	0.0723
	EM	0.0942	0.1062	0.0416	0.0652	0.0276	0.0729
$\hat{\gamma}_1$	MPL	0.0483	0.0252	0.0051	0.0104	0.0125	0.0078
	EM	0.0145	0.0035	0.0037	0.0066	0.0118	0.0071
π^R		0.75	0.64	0.75	0.64	0.75	0.64
$\hat{\beta}_0$	MPL	-0.2806	-0.3503	-0.2838	-0.3033	-0.3011	-0.3267
	EM	-0.8087	-0.7592	-0.7683	-0.8035	-0.7786	-0.7992
$\hat{\beta}_1$	MPL	0.0192	-0.1575	-0.0070	-0.0368	-0.0394	-0.0543
	EM	0.0301	-0.1309	-0.0378	-0.0150	-0.0391	-0.0537
$\hat{\gamma}_1$	MPL	0.0103	0.0720	0.0397	-0.0004	-0.0081	-0.0072
	EM	-0.0235	0.0570	0.0053	-0.0180	-0.0093	-0.0072

Note that the values of π^R presented here are the proportions of right censored observations in the whole sample i.e. including the right censored observations from individuals in the cured fraction. Scenarios with π^R of 0.59 or 0.41 correspond to scenarios with the censoring distribution parameter $\lambda_c = 4.2$. Scenarios with π^R of 0.75 or 0.64 correspond to scenarios with $\lambda_c = 1.25$.

the standard error asymptotically compared to obtaining it via bootstrapping. For β_1, both methods tend to have fair agreement between the estimated and Monte Carlo standard errors in the two larger samples sizes of $n = 500$ and $n = 2000$, although the performance is impacted by increases in the right censored proportion. In the smallest sample size of $n = 100$, there is significant disagreement between the estimated and Monte Carlo standard errors for $\hat{\beta}_1$ for both methods. However, the differences between the two are generally greater for the bootstrapped (EM) standard errors compared to the MPL asymptotic estimates. Additionally, bootstrapping has produced a number of extremely large standard error estimates while the MPL asymptotic estimation method has not. 95% coverage probabilities (see Appendix B2) for both methods are reasonable in the two larger sample sizes. Table 3 also displays the estimated and Monte Carlo standard errors for γ_1. Evidently, when right censoring is lower, both methods perform well with good agreement between the estimated and Monte Carlo standard errors for this parameter. As right censoring increases, both methods have a tendency to overestimate the standard error of $\hat{\gamma}_1$ in the larger samples. This is reflected in the overly large 95% coverage probabilities for

Table 3: Mean estimated SE and (Monte Carlo) SE for estimates of β_1 and γ_1

n		100	100	500	500	2000	2000
$\pi(\mathbf{z})$		0.5	0.75	0.5	0.75	0.5	0.75
π^R		0.59	0.41	0.59	0.41	0.59	0.41
$\hat{\beta}_1$	MPL	0.5229	0.6828	0.2302	0.2826	0.1146	0.1395
		(0.5392)	(0.6948)	(0.2451)	(0.2900)	(0.1360)	(0.1246)
	EM	0.6597	2.3844	0.2316	0.2899	0.1155	0.1408
		(0.5349)	(0.6942)	(0.2452)	(0.2904)	(0.1362)	(0.1248)
$\hat{\gamma}_1$	MPL	0.3586	0.2939	0.1503	0.1263	0.0764	0.0630
		(0.3944)	(0.2982)	(0.1534)	(0.1392)	(0.0800)	(0.0657)
	EM	0.3991	0.3013	0.1572	0.1284	0.0776	0.0629
		(0.3950)	(0.2972)	(0.1521)	(0.1392)	(0.0797)	(0.0659)
π^R		0.75	0.64	0.75	0.64	0.75	0.64
$\hat{\beta}_1$	MPL	0.5829	0.6382	0.2514	0.2840	0.1249	0.1404
		(0.6232)	(0.5831)	(0.2818)	(0.2336)	(0.1266)	(0.1463)
	EM	0.7059	1.0503	0.2551	0.2939	0.1242	0.1403
		(0.6150)	(0.6174)	(0.2476)	(0.3079)	(0.1267)	(0.1466)
$\hat{\gamma}_1$	MPL	0.4654	0.3765	0.2010	0.1635	0.0983	0.0811
		(0.5689)	(0.4073)	(0.2019)	(0.1699)	(0.0979)	(0.0745)
	EM	0.7626	0.4492	0.2038	0.1667	0.1005	0.0811
		(0.5043)	(0.4078)	(0.1810)	(0.1625)	(0.0981)	(0.0744)

The mean estimated standard error reported for the MPL method is the asymptotic estimate, and the mean estimated standard error reported for the EM method is a bootstrapped estimate.

these scenarios (see Appendix B2). In the smallest sample size, the bootstrapped estimates of the standard error are similarly overestimated, but the asymptotic MPL estimates of the standard error tend to be underestimated. Accordingly, the 95% coverage probabilities for the MPL method are slightly low. Overall, the performance of the MPL method for estimating β_1 and γ_1 is competitive with the performance of the existing method, and the asymptotic standard error estimator performs as well as bootstrapping for these parameters.

4.1.2 Cumulative baseline hazard function estimation results

This simulation study also compared the estimation of the cumulative baseline hazard function, $H_0(t)$, in both methods considered here. The cumulative baseline hazard function was chosen as a point of comparison because the `smcure` package does not offer any direct estimate of the baseline hazard function. However, the estimated baseline survival function, $S_0(t)$, is available. Obtaining the cumulative baseline hazard function estimate from the `smcure` package

is therefore a simple matter of taking $H_0(t) = -\ln(S_0(t))$. Obtaining the MPL estimate of the approximated cumulative baseline hazard function is similarly straightforward. Recall that the baseline hazard function is approximated by $h_0(t) = \boldsymbol{\psi}(t)\boldsymbol{\theta}$, where $\boldsymbol{\psi}(t)$ is a set of cubic M-spline basis functions and $\boldsymbol{\theta}$ is a parameter vector of length m. As noted in Section 2.1.1, a convenient feature of using M-spline basis functions for approximating the baseline hazard function is that computation of the cumulative baseline hazard function is simple. Each M-spline $\boldsymbol{\psi}_u(t)$ can be associated with an I-spline, its integral, denoted as $\boldsymbol{\Psi}_u(t)$. Each $\boldsymbol{\Psi}_u(t)$ will be monotonically increasing when associated with the coefficient vector $\boldsymbol{\theta}$. An appropriate approximation to the cumulative baseline hazard function can therefore be found by $H_0(t) = \boldsymbol{\Psi}(t)\boldsymbol{\theta}$. In order to evaluate the estimates of $H_0(t)$, the bias of the estimate was considered at three time values, t_1, t_2 and t_3, which respectively correspond to the first, second, and third quartile of T.

The `smcure` package does not provide either asymptotic or bootstrapped estimates of the standard error of the baseline survival or hazard functions. As such, for the purposes of comparison, only the Monte Carlo standard errors of the two methods will be computed and discussed here. However, it should be noted the asymptotic standard error of the MPL estimate for the cumulative baseline hazard function can be easily obtained using the delta method. The use of the delta method for estimating the standard error of the baseline hazard function is explored in the second simulation study (see Section 4.2.2). The ability to make large sample inferences on survival quantities can be considered a key strength of the proposed MPL method when compared to existing alternatives.

Table 4 summarises the bias in the two estimates of the cumulative baseline hazard function at t_1, t_2 and t_3. It is clear that the MPL estimate outperforms the Breslow (EM) estimate across all simulation scenarios. In both methods, there is an obvious trend of bias increasing with time t. However, the magnitude of this increase is consistently far smaller for the MPL estimate. Notably, the bias in the MPL estimate does not appear to be strongly impacted by small sample sizes at earlier values of t, while the bias in the Breslow (EM) estimate increases in smaller samples. Table 4 also shows the Monte Carlo standard errors for each simulation scenario for both of the methods considered. These results allow for a comparison of the variability in the two different estimates of the cumulative baseline hazard function. It is evident that in the larger sample sizes, $n = 500$ and $n = 2000$, the two methods are largely comparable, especially at earlier values of t. In smaller samples with $n = 100$, it appears that the MPL estimate tends to be less variable than the Breslow (EM) estimate. Overall, it is clear that the MPL method of estimating the cumulative baseline hazard function is superior to the Breslow (EM) method.

Table 4: Bias and Monte Carlo standard error for estimates of $H_0(t)$

Bias in estimate for $H_0(t)$							
	n	100	100	500	500	2000	2000
	$\pi(\mathbf{z})$	0.5	0.75	0.5	0.75	0.5	0.75
	π^R	0.59	0.41	0.59	0.41	0.59	0.41
t_1	MPL	0.0153	0.0356	0.0258	0.0126	0.0203	0.0158
	EM	0.1026	0.1113	0.0853	0.0745	0.0476	0.0768
t_2	MPL	0.0327	0.0688	0.0465	0.0159	0.0432	0.0325
	EM	0.1885	0.2422	0.1894	0.1709	0.1176	0.1819
t_3	MPL	0.0406	0.0744	0.0840	0.0305	0.0710	0.0553
	EM	0.3229	0.4389	0.3909	0.3544	0.2328	0.3641
	π^R	0.75	0.64	0.75	0.64	0.75	0.64
t_1	MPL	-0.0176	-0.0273	-0.0213	-0.0189	-0.0234	-0.0195
	EM	0.0485	0.0500	0.0427	0.0501	0.0436	0.0455
t_2	MPL	-0.0458	-0.0430	-0.0387	-0.0392	-0.0483	-0.0393
	EM	0.0905	0.1141	0.1156	0.1251	0.1083	0.1180
t_3	MPL	-0.1762	-0.1276	-0.0923	-0.0698	-0.0965	-0.0685
	EM	0.2054	0.2682	0.2306	0.2586	0.2359	0.2548
Monte Carlo standard error for estimate of $H_0(t)$							
	n	100	100	500	500	2000	2000
	$\pi(\mathbf{z})$	0.5	0.75	0.5	0.75	0.5	0.75
	π^R	0.59	0.41	0.59	0.41	0.59	0.41
t_1	MPL	0.1137	0.0794	0.0442	0.0311	0.0172	0.0142
	EM	0.1057	0.1012	0.0427	0.0352	0.0477	0.0170
t_2	MPL	0.1867	0.1466	0.0785	0.0539	0.0327	0.0287
	EM	0.1765	0.2061	0.0790	0.0628	0.1140	0.0337
t_3	MPL	0.3212	0.2787	0.1433	0.1098	0.0615	0.0454
	EM	0.3534	0.3753	0.1388	0.1330	0.2267	0.0532
	π^R	0.75	0.64	0.75	0.64	0.75	0.64
t_1	MPL	0.1132	0.0874	0.0511	0.0394	0.0229	0.0157
	EM	0.1266	0.0843	0.0479	0.0480	0.0249	0.0200
t_2	MPL	0.2255	0.1713	0.1006	0.0842	0.0390	0.0327
	EM	0.2752	0.2015	0.0946	0.0887	0.0517	0.0420
t_3	MPL	0.4918	0.3524	0.1856	0.1644	0.0844	0.0704
	EM	0.4828	0.3799	0.1694	0.1656	0.1058	0.0857

4.2 Simulation study using partly-interval censored data

The aim of the second simulation study was to assess the performance of the proposed method of fitting a proportional hazards mixture cure model to partly-interval censored survival data. This was not a comparative study because there is, at present, a very limited number of comparable methods available. The performance of the regression parameter estimates $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and the asymptotic variance estimator for these parameters, was investigated (Section 4.2.1). Additionally, the estimate of the baseline hazard function and its asymptotic standard error estimator were evaluated (Section 4.2.2).

Table 5 outlines the conditions considered as part of this simulation study. Where possible, the range of conditions considered are similar to those used in the first simulation study. As in the previous simulation study, this study considers scenarios with two cured fraction sizes and uses a single binomial covariate, z_1, in the logistic regression for the cure probability. As above, the size of the cured fraction is controlled by the true value of β_0, the intercept in the logistic regression model. Recall that a value of $\beta_0 = 1$ corresponds to a non-cured proportion of approximately 0.75 in the group with covariate $z_1 = 0$ and a value of $\beta_0 = 0$ corresponds to a proportion of approximately 0.5. Prior to obtaining any event or censoring times, values of U_i^C, $\pi(\boldsymbol{z}_i)$, and u_i were generated for each individual i using the same method outlined previously. Again, observed times for the cured fraction, all of which constituted right censoring times, were drawn from an exponential distribution.

For individuals in the non-cured fraction, where $u_i = 1$, event times were drawn from a Weibull distribution. The observed survival times for these individuals were made up of event times, right censoring times, left censoring times, and interval censoring times. The observed survival times (T_i^L, T_i^R) for individuals in the non-cured fraction, including both event and censoring times, were obtained by

$$T_i^L = Y_i^{\delta(U_i^E<\pi^E)}(\alpha_L U_i^L)^{\delta(\pi^E\leq U_i^E,\alpha_L U_i^L\leq Y_i\leq\alpha_R U_i^R)}(\alpha_R U_i^R)^{\delta(\pi^E\leq U_i^E,\alpha_R U_i^R<Y_i)}0^{\delta(\pi^E\leq U_i^E,Y_i<\alpha_L U_i^L)}$$

$$T_i^R = Y_i^{\delta(U_i^E<\pi^E)}(\alpha_L U_i^L)^{\delta(\pi^E\leq U_i^E,Y_i<\alpha_L U_i^L)}(\alpha_R U_i^R)^{\delta(\pi^E\leq U_i^E,\alpha_L U_i^L\leq Y_i\leq\alpha_R U_i^R)}\infty^{\delta(\pi^E\leq U_i^E,\alpha_R U_i^R<Y_i)}$$

where Y_i denotes the event time drawn from a Weibull distribution, π^E denotes the event probability in the non-cured fraction, U_i^L, U_i^R and U_i^E denote independent standard uniform variables, and α_L and α_R are scalars that define interval censoring values. Note that $\delta(\cdot)$ denotes an indicator function, and that here $0^0 = \infty^0 = 1$. A single binomial covariate, x_1, independent of the covariate used in the logistic regression, was generated for the proportional hazards regression.

In this simulation study, two different values for the event probability in the non-cured fraction were considered, as shown in Table 5. There were unbalanced proportions of left, right and interval censoring. Out of the censored observations in each non-cured fraction, 66.7% were right censored, 9.5% were left censored, and 23.8% were interval censored. However, these specifications did not correspond to the rate of censoring in the sample as a whole, because

Table 5: Specifications for Simulation Study 2

Simulation Parameters	
$\boldsymbol{\gamma}$ vector	$[0.5]$
$\mathbf{X}$ vector	$\mathbf{X} = [x_1]$
Y distribution	Weibull
Baseline hazard	$h_0(y) = 3y^2$
α_L and α_R	$(0.9, 1.3)$
Cured Fraction Scenarios	
$\boldsymbol{\beta}$ vector	$[1, 1]^T$ giving $\pi(\mathbf{z}) \simeq 0.75$ $[0, 1]^T$ giving $\pi(\mathbf{z}) \simeq 0.5$
$\mathbf{Z}$ vector	$\mathbf{Z} = [1, z_1]$
Simulation Scenarios	
Sample size	$n = 100, 500, 2000$
Percentage of events (in non-cured fraction)	$25\%, 50\%$

Note that the size of the cured fraction and the censoring rate in the non-cured fraction contribute to the rate of censoring in the sample as a whole.

right censored times were recorded for all individuals in the cured fraction. As discussed above, this means that both the event probability and the size of the cured fraction effect the extent of censoring in the sample. In some cases, the combination of these two parameters produced samples with extremely high levels of right censoring. In the results that follow, the values presented for π^E and π^R are the proportion of event times and right censoring times in the sample as a whole.

As above, cubic M-splines with some number n_κ knots were used to approximate the baseline hazard function. The number of knots was selected using the same process laid out in the previous simulation study. The minimum number of knots used for any simulation scenario was again $n_\kappa = 3$. The number of knots selected for each simulation scenario can be found in Appendix C1. As above, the value of the smoothing parameter λ was set manually to zero.

4.2.1 Regression parameter estimation results

The biases in the estimates of the parameters β_0, β_1 and γ_1 for all simulation scenarios are presented in Table 6. These results indicate that the proposed model performs well when there is a reasonably large proportion of events in the sample. Specifically, when π^E is 0.25 or 0.36, the bias in the estimates of all three parameters is fairly low, particularly when the sample size is larger. The bias in the estimates of β_1 and γ_1 generally remains low when the proportion of events decrease. In particular, the estimate of γ_1 is fairly robust to increases in right censoring and increases in the size of the cured fraction, especially when the sample size is large.

Table 6: Bias in estimates of β_0, β_1 and γ_1

n	100	100	500	500	2000	2000
$\pi(\mathbf{z})$	0.5	0.75	0.5	0.75	0.5	0.75
π^E	0.25	0.36	0.25	0.36	0.25	0.36
π^R	0.67	0.52	0.67	0.52	0.67	0.52
$\hat{\beta}_0$	0.0575	0.0442	0.0409	0.0382	0.0256	0.0318
$\hat{\beta}_1$	0.0897	0.0586	0.0366	-0.0095	0.0059	0.0220
$\hat{\gamma}_1$	0.0515	0.0591	0.0148	0.0265	0.0152	0.0235
π^E	0.13	0.18	0.13	0.18	0.13	0.18
π^R	0.75	0.64	0.75	0.64	0.75	0.64
$\hat{\beta}_0$	-0.4149	-0.3316	-0.3488	-0.4148	-0.3655	-0.4200
$\hat{\beta}_1$	0.0082	-0.0275	-0.0278	-0.0647	-0.0720	-0.1028
$\hat{\gamma}_1$	0.0546	0.0284	0.0419	0.0371	-0.0186	0.0144

Note that the values of π^E and π^R presented here are the proportions of events and right censored observations in the whole sample i.e. including right censored observations from individuals in the cured fraction. Scenarios with a π^E of 0.25 or 0.36 and a π^R of 0.67 or 0.52 correspond to an event probability in the non-cured fraction of 50%. Scenarios with a π^E of 0.13 or 0.18 and a π^R of 0.75 or 0.64 correspond to an event probability in the non-cured fraction of 25%.

However, the same cannot be said for the estimate of β_0. Clearly, when the proportion of events decreases, the bias in the estimate of β_0 increases significantly in magnitude. This increased bias is consistently negative, which can be interpreted as a systematic overestimation of the cured fraction. It is important to note the role that the size of the cured fraction plays in causing the increased bias in the estimate of β_0. Evidently, the proportion of right censoring in the sample is not the only driving factor behind this increased bias. In fact, Table 6 shows that the bias in β_0 can increase even as the overall proportion of right censoring in the sample π^R decreases, if there is a change in the size of the cured fraction. This is obvious when comparing the scenarios with $\pi(\mathbf{z}) = 0.5$ and $\pi^R = 0.67$ with the scenarios with $\pi(\mathbf{z}) = 0.75$ and $\pi^R = 0.64$ across all three sample sizes. Here, it is not the amount of right censoring in the sample itself that is producing larger bias in the estimate of β_0. Instead, it is the fact that a decreased proportion of these right censored observations belong to the cured fraction, as the size of the non-cured fraction $\pi(\mathbf{z})$ increases. This produces a larger number of non-cured right censoring times which could potentially be misidentified as part of the cured fraction. In turn, this leads to an overestimation of the size of the cured fraction. This may indicate that there is poor separation between the cured and uncured sub-populations. There is some evidence that this phenomena also leads to increasing bias in the estimate of β_1 when event probabilities

Table 7: Mean asymptotic and (Monte Carlo) standard errors of of β_0, β_1 and γ_1

n	100	100	500	500	2000	2000
$\pi(\mathbf{z})$	0.5	0.75	0.5	0.75	0.5	0.75
π^E	0.25	0.36	0.25	0.36	0.25	0.36
π^R	0.67	0.52	0.67	0.52	0.67	0.52
$\hat{\beta}_0$	0.2798	0.3320	0.1207	0.1429	0.0599	0.0708
	(0.3212)	(0.3292)	(0.1204)	(0.1361)	(0.0588)	(0.0672)
$\hat{\beta}_1$	0.5564	0.6676	0.2383	0.2819	0.1182	0.1400
	(0.6273)	(0.7377)	(0.2412)	(0.2653)	(0.1204)	(0.1509)
$\hat{\gamma}_1$	0.4443	0.3614	0.1900	0.1562	0.0943	0.0773
	(0.4669)	(0.3587)	(0.1618)	(0.1579)	(0.0881)	(0.0717)
π^E	0.13	0.18	0.13	0.18	0.13	0.18
π^R	0.75	0.64	0.75	0.64	0.75	0.64
$\hat{\beta}_0$	0.2952	0.3504	0.1287	0.1439	0.0637	0.0712
	(0.2689)	(0.3694)	(0.1326)	(0.1528)	(0.0575)	(0.0684)
$\hat{\beta}_1$	0.6436	0.6208	0.2712	0.2651	0.1246	0.1385
	(0.5864)	(0.7309)	(0.2501)	(0.2849)	(0.1204)	(0.1376)
$\hat{\gamma}_1$	0.6394	0.5092	0.2530	0.2058	0.1218	0.1012
	(0.5303)	(0.4504)	(0.2204)	(0.1849)	(0.1095)	(0.0876)

are low. Overall, the results in Table 6 indicate that the model performs well when there is a reasonable number of observed event times in the sample, and particularly when the sample size is large. The model may be sensitive to changes in the cured fraction and poorly separated sub-populations, but the proportional hazards regression parameter estimate γ_1 consistently performs well.

Table 7 presents the mean asymptotic and the Monte Carlo estimates of the standard error for each regression parameter, allowing for the evaluation of the MPL asymptotic estimate. The 95% coverage probabilities for the regression parameters, found in Appendix C2, offer further insight into the performance of the standard error estimation process. Table 7 indicates that, for β_0, there is generally good agreement between the estimated asymptotic and Monte Carlo standard errors across the majority of the different scenarios considered, with the only exception being in the cases of the smallest sample size $n = 100$. There does not appear to be any particularly strong effect on the performance of the asymptotic standard error estimation due to changes in the size of the cured fraction. However, the 95% coverage probabilities for β_0 are extremely low for scenarios with decreased event probabilities of $\pi^E = 0.13$ or $\pi^E = 0.18$. This is a result of the bias in the estimation of this parameter, discussed above, rather than a reflection of the quality of the standard error estimate. There is similarly good agreement between the

estimated asymptotic and Monte Carlo standard errors for the parameter β_1. Again, the only exception to this is scenarios with the smallest sample size. Decreasing the event probability π^E does little to diminish performance, particularly in larger samples. Appendix C2 also shows that 95% coverage probabilities for β_1 are reasonable in most cases. The only exception are the scenarios with lower event probabilities π^E and $n = 2000$. This is likely due to the increased bias in the estimates of β_1 in these scenarios, seen in Table 6.

In samples with higher event probabilities, there is again reasonable agreement between the estimated asymptotic and Monte Carlo standard errors of γ_1. However, it is evident in Table 7 that the discrepancy between the two increases with decreasing event probabilities. Notably, there is a persistent tendency for the asymptotic standard error estimate to be larger than its equivalent Monte Carlo standard error when the event probability is low. This indicates that the asymptotic standard error is being systematically overestimated under these conditions. This is likely related to the systematic overestimation of the cured fraction, discussed above. The overestimation of the cured fraction reduces the share of the sample on which the proportional hazards model estimation is based and thus produces larger standard error estimates. Given this, it is unsurprising that the 95% coverage probabilities for γ_1 in samples with high levels of right censoring are generally too high. Nonetheless, the 95% coverage probabilities for γ_1 in samples with higher event probabilities are generally appropriate. Overall, the results in Table 7 demonstrate that the asymptotic standard error estimator in this model generally performs well. Where the performance of this estimator is poorer, it is usually a result of bias in the parameter estimates of some elements of the model under certain conditions.

4.2.2 Baseline hazard function estimation results

This simulation study also evaluated the MPL estimates of the baseline hazard function $h_0(t)$ for partly-interval censored data. In order to do so, the bias of the baseline hazard function estimate was considered at three time values, t_1, t_2 and t_3, which respectively correspond to the first, second, and third quartile of T. The agreement between the estimated asymptotic and Monte Carlo standard error at these three values of t was also considered. Note that the asymptotic standard error of the baseline hazard function was estimated using the delta method. Recall that $h_0(t) = \boldsymbol{\psi}(t)\boldsymbol{\theta}$, meaning that the baseline hazard function can be treated as a function of $\boldsymbol{\theta}$. As such, the variance of the baseline hazard function estimate can be estimated by finding

$$Var(h_0(t)) = \left(\frac{\partial h_0(t)}{\partial \theta_u}\right)^T Cov(\boldsymbol{\theta}) \frac{\partial h_0(t)}{\partial \theta_u} = \boldsymbol{\psi}(t)^T Cov(\boldsymbol{\theta}) \boldsymbol{\psi}(t)$$

This estimated standard error of the baseline hazard function was also used to produce 95% coverage probabilities at the three values of t mentioned above.

Table 8 summarises the bias in the estimates of the baseline hazard function at t_1, t_2, and t_3. Firstly, it is clear that in the vast majority of scenarios considered, the bias increases with

the value of t. This is unsurprising given that the number of observations generally lessens over time. In scenarios with a higher event probability, the bias is generally reasonable at t_1 and t_2. In scenarios with $n = 500$ or $n = 2000$, the extent of the increase in the bias at later values of t is not as large in scenarios with $\pi^E = 0.25$ and the smaller non-cured fraction size of $\pi(\mathbf{z}) = 0.5$ as it is in scenarios with a larger event probability of $\pi^E = 0.36$ and a larger non-cured fraction of $\pi(\mathbf{z}) = 0.75$. This may indicate that the size of the cured fraction has some influence on the performance of the baseline hazard function estimate, especially at later values of t. When event probabilities are low, the bias in the estimate becomes large across all values of t, although the aforementioned trend over time is still evident. Generally, the estimate of the baseline hazard function performs well at earlier values of t in samples with a reasonable proportion of event times.

The mean estimated asymptotic and Monte Carlo standard errors for the baseline hazard function at t_1, t_2 and t_3 are also shown in Table 8. Corresponding 95% coverage probabilities can be found in Appendix C3. Generally there is good agreement between the asymptotic and Monte Carlo standard errors across values of t for larger samples with higher event probabilities. Accordingly, the 95% coverage probabilities for these scenarios are also appropriate. Discrepancy between the asymptotic and Monte Carlo standard errors increases when the sample size is smaller due to the lower availability of observations. In scenarios with lower event probabilities, there is generally still good agreement between the asymptotic and Monte Carlo standard errors when the sample size is relatively large, especially at earlier values of t. Generally for scenarios with lower event probabilities, 95% coverage probabilities are inappropriate, likely due to a combination of large bias and poor standard error estimation. Generally, the standard error estimation process for the baseline hazard function performs well for relatively large samples, at earlier values of t, and when levels of censoring are not extreme.

Table 8: Bias, estimated standard errors and (Monte Carlo) standard errors for $h_0(t)$

Bias in estimates for $h_0(t)$						
n	100	100	500	500	2000	2000
$\pi(\mathbf{z})$	0.5	0.75	0.5	0.75	0.5	0.75
π^E	0.25	0.36	0.25	0.36	0.25	0.36
π^R	0.67	0.52	0.67	0.52	0.67	0.52
t_1	0.0433	-0.0185	-0.0053	-0.0573	-0.0073	-0.0539
t_2	-0.0408	-0.1868	-0.0117	-0.0484	-0.0275	-0.0349
t_3	-0.3733	-0.3496	-0.0267	0.1014	-0.0226	0.0813
π^E	0.13	0.18	0.13	0.18	0.13	0.18
π^R	0.75	0.64	0.75	0.64	0.75	0.64
t_1	-0.1853	-0.1616	-0.2557	-0.2840	-0.2447	-0.1788
t_2	-0.7353	-0.5274	-0.4797	-0.4425	-0.4170	-0.2828
t_3	-1.7197	-1.4448	-0.8085	-0.3711	-0.4165	-0.3405
Mean asymptotic and (Monte Carlo) standard error estimates						
n	100	100	500	500	2000	2000
$\pi(\mathbf{z})$	0.5	0.75	0.5	0.75	0.5	0.75
π^E	0.25	0.36	0.25	0.36	0.25	0.36
π^R	0.67	0.52	0.67	0.52	0.67	0.52
t_1	0.2621	0.3559	0.1551	0.2075	0.0776	0.1038
	(0.3435)	(0.3688)	(0.1590)	(0.2039)	(0.0815)	(0.0962)
t_2	0.4135	0.6839	0.3186	0.3352	0.1614	0.1612
	(0.5714)	(0.6655)	(0.3059)	(0.3297)	(0.1596)	(0.1689)
t_3	1.1892	1.3733	0.5284	0.6327	0.2554	0.2978
	(1.5261)	(1.3310)	(0.5250)	(0.4942)	(0.2528)	(0.2978)
π^E	0.13	0.18	0.13	0.18	0.13	0.18
π^R	0.75	0.64	0.75	0.64	0.75	0.64
t_1	0.4588	0.4407	0.2262	0.2628	0.0902	0.0841
	(0.4124)	(0.4266)	(0.1933)	(0.2203)	(0.0734)	(0.0689)
t_2	1.0764	0.9923	0.3231	0.4928	0.1655	0.1433
	(1.2248)	(1.0460)	(0.3243)	(0.4001)	(0.1454)	(0.1428)
t_3	2.9201	2.7437	0.8786	0.8745	0.2779	0.3250
	(3.0374)	(2.2711)	(0.7874)	(0.8436)	(0.2248)	(0.3566)

5 R Package and Implementation

This chapter will present an overview of the set of R functions developed to fit the proposed method. It will also offer an example of how this method could be used to analyse a real dataset. Access to the relevant code is available online (see Appendix D). The structure of the package presented here closely follows that of the package `survivalMPL` (Couturier et al. 2017), which implements the method laid out in Ma et al. (2019).

5.1 Package description

The estimation methods presented in this thesis are implemented in the `phmc_mpl` package. The model can be fitted in R using the code `phmc_mpl(phformula, piformula, data, control, ...)`. The required arguments are as follows:

- `phformula`: a formula object, with the response on the righthand side of a $\sim$ operator and the covariates for the proportional hazards regression for the latency on the lefthand side. The response must be in the form of a survival object created using the `Surv` function from the `survival` package.
- `piformula`: a specification of the covariates to be used in the logistic regression for the incidence, on the lefthand side of a $\sim$ operator. Specifying ~ 1 will produce an estimate of the intercept only.
- `data`: a data frame containing the covariates used in the `phformula` and `piformula` arguments.
- `control`: an object of class `phmc_mpl.control`, specifying control options such as number of knots, maximum number of iterations, and the convergence limit.

The full range of options available for specification as part of the `phmc_mpl.control` function is as follows:

- `maxIter=c(150,75e+02,1e+04)`: a vector of three integers. The first is the maximum number of iterations allowed for the smoothing parameter estimation. The second is the maximum number of iterations allowed for the estimation of the regression parameters and baseline hazard function. The third is the total number of iterations allowed.
- `epsilon=c(1e-16,1e-10)`: a vector of two values indicating the minimum distance of, respectively, the survival function from 1 and the baseline hazard function from 0, specified in order to prevent problems with calculating logarithms in the algorithm.
- `kappa=1/0.6`: a value greater than 1, used in the algorithm to decrease the step size if the penalised likelihood does not increase after a given iteration.

- `conv_limit=1e-4`: a convergence tolerance value, defining the minimum absolute difference between the parameter estimates at consecutive iterations that must be reached for convergence to be achieved.

- `smooth=NULL`: the value of the smoothing parameter λ. When left as `NULL`, the smoothing parameter is estimated as outlined in Section 3.4.1. When specified manually, it should be greater than or equal to 0.

- `min.theta=1e-10`: a value indicating the minimum possible value for any the baseline hazard parameter, θ_u. Any value of θ_u less than `min.theta` will be considered zero.

- `n.knots=8`: an integer specifying the number of internal quantile knots to be used. Note that the total number of knots n_κ will always be two greater than the number specified here, as n_κ include the two external knots placed at the minimum and maximum observations.

- `order=3`: an integer specifying the order of M-splines to be used.

- `range.quant=c(0.075,0.9)`: a vector of two values, defining the range of the quantile knots. Under the default values, `n.knots` quantile knots are set between the quantiles 0.075 and 0.9 of the observed event times.

- `ties="epsilon"`: a character string specifying how to handle duplicated outcomes when defining the knot sequence. The default option `"epsilon"` adds random noise smaller than 1e-10 to each duplicated observed survival time. An alternative option is `"unique"` which deletes duplicated fully survival times.

- `seed=NULL`: `NULL` or an integer vector compatible with `.Random.seed` to be used when adding random noise to duplicate events when `ties="epsilon"`

Other functions available in the package are:

- `plot.phmc_mpl`: plots the smooth estimate of the baseline hazard function, cumulative baseline hazard function, and/or baseline survival function with 95% confidence interval.

- `predict.phmc_mpl`: predicts the hazard function or the survival function, standard error, and 95% confidence interval using the estimated model and a given set of covariate values.

5.2 Example of application to real-life data

To illustrate the use of the `R` implementation detailed above, this section will fit a proportional hazards mixture cure model to a real dataset. The dataset used is the `e1684` data found in the `smcure` package, sourced from the study done by Kirkwood et al. (1996). This study evaluated the use of interferon alfa-2b as an adjuvant therapy following surgery on melanoma.

The dataset contains three covariates, treatment (0 = control, 1 = treatment), gender (0 = male, 1 = female), and age (as a continous mean-centered variable). Note that this dataset only contains observed and right censored survival times. After the removal of missing values, the dataset has a sample size of 284. The creators of the `smcure` package also used this dataset to illustrate the functionality of that package for fitting a proportional hazards mixture cure model, detailed in Cai et al. (2012*a*).

The mixture cure proportional hazards model can be fitted as follows. Note that this code uses the default number of internal quantile knots. Selection of the smoothing parameter will be performed automatically using the method outlined in Section 3.4.1.

```
> e.surv=Surv(time=e1684$FAILTIME,event=e1684$FAILCENS)
> model1=phmc_mpl(e.surv~e1684$TRT+e1684$AGE+e1684$SEX,
+          pi.formula=~e1684$TRT+e1684$AGE+e1684$SEX,data=e1684,
+          phmc_mpl.control(n.knots=c(8),smooth=NULL))
```

This code produces the following output.

```
> summary(model1)

phmc_mpl(ph.formula = e.surv ~ e1684$TRT + e1684$SEX + e1684$AGE,
    pi.formula = ~e1684$TRT + e1684$SEX + e1684$AGE, data = e1684,
    control = phmc_mpl.control(n.knots = c(8), smooth = NULL))

-----
Mixture Cure Proportional Hazards Model Fitted Using MPL
Penalized log-likelihood  :  -164.5312
Estimated smoothing value :  4.743324
Convergence : Yes, 715 iterations

-----
Logistic regression parameters :
           Estimate Std. Error z-value Pr(>|z|)
           2.112622   1.428859  1.4785   0.1393
e1684$TRT -0.234118   0.850133 -0.2754   0.7830
e1684$SEX -0.397047   0.830328 -0.4782   0.6325
e1684$AGE  0.054493   0.057087  0.9546   0.3398

Proportional hazards regression parameters :
            Estimate Std. Error z-value Pr(>|z|)
e1684$TRT -0.3889650  0.2372908 -1.6392   0.1012
```

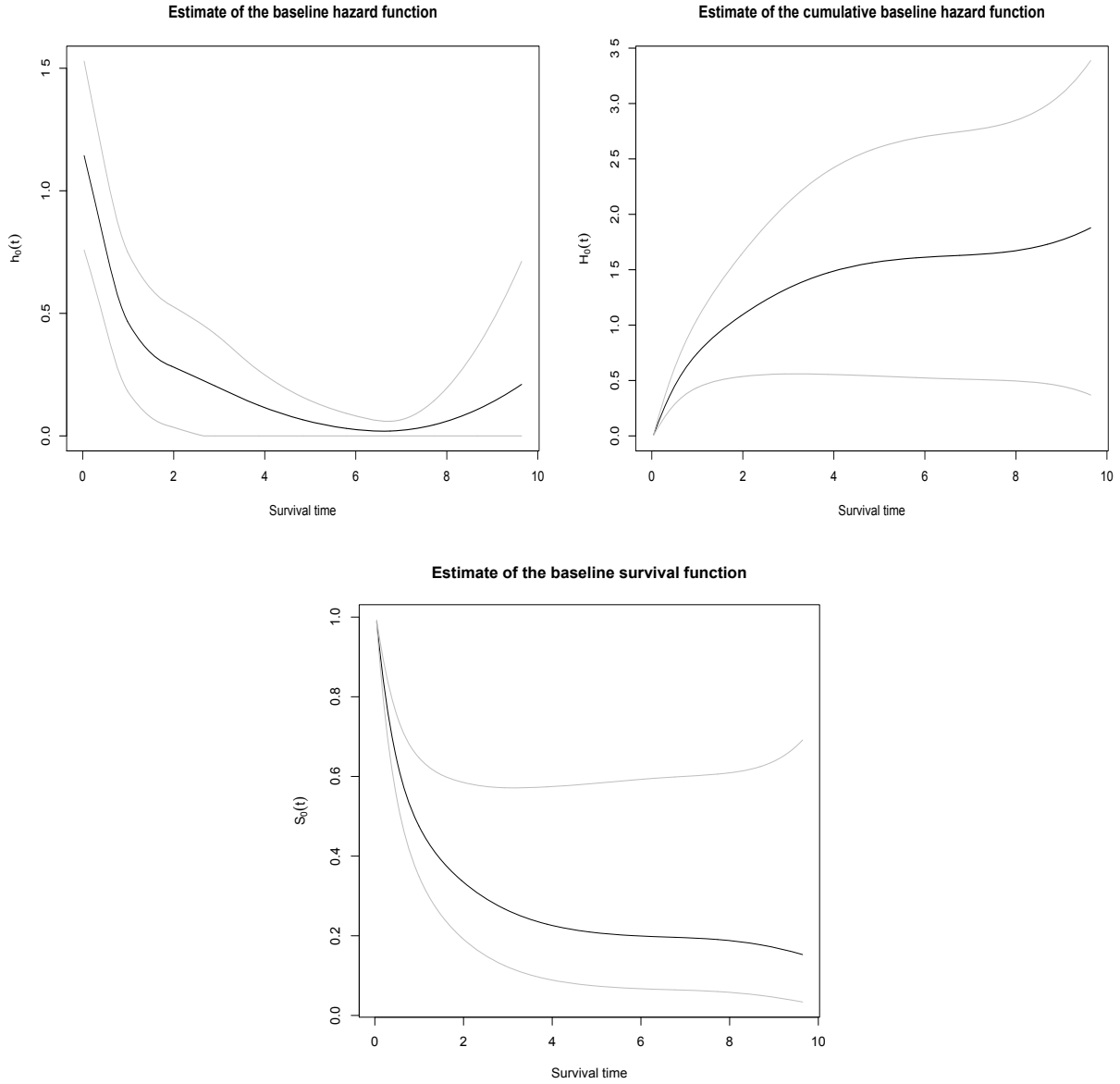
```
e1684$SEX  0.1049943  0.2483536  0.4228   0.6725
e1684$AGE -0.0073339  0.0088921 -0.8248   0.4095

-----
```

Available in the output are the logistic regression parameter estimates for the intercept and the three covariates used to estimate the incidence, and the proportional hazards regression parameter estimates for the three covariates used to model the latency. The asympotic variance estimates for all parameters are also available, as well as a test of significance. The output also reports that convergence was achieved after 715 iterations in total, and that the value of the smoothing parameter λ automatically selected was approximately 4.74.

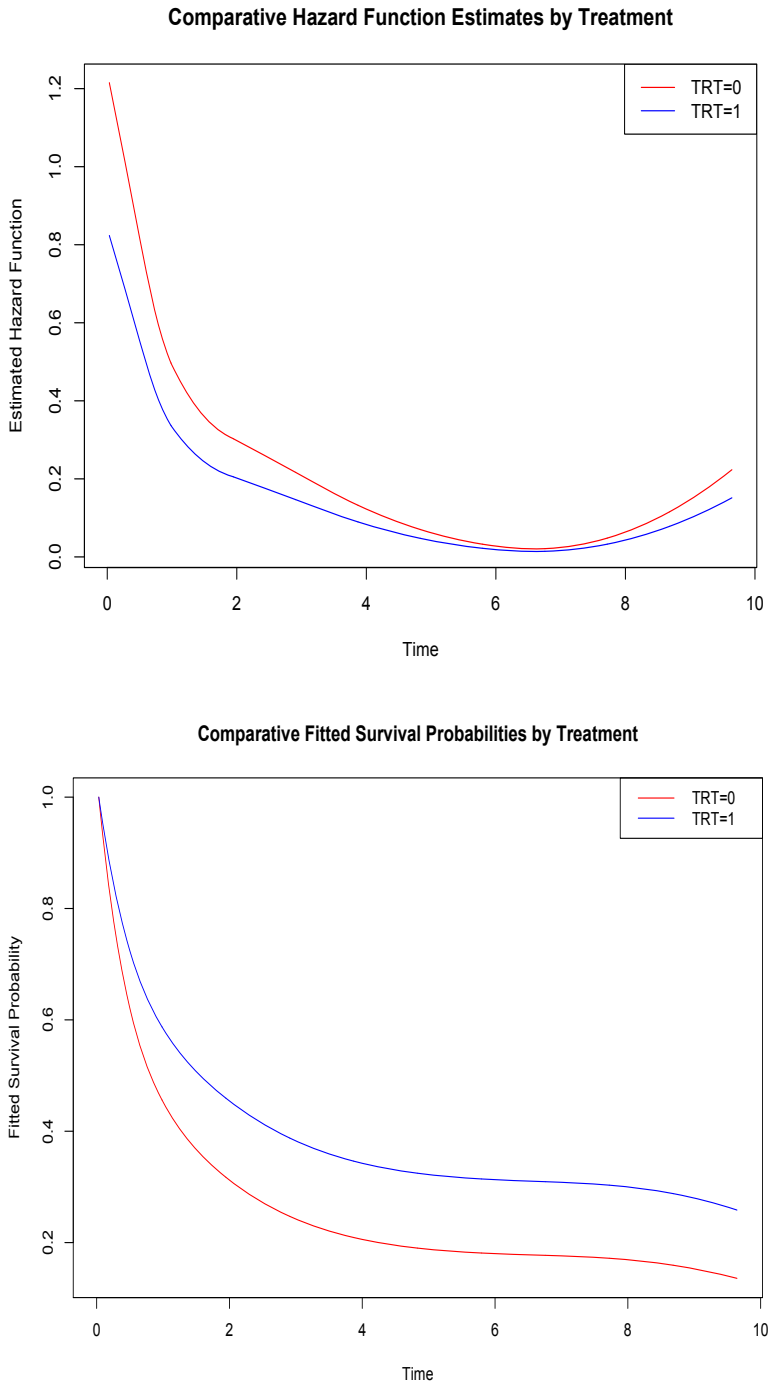
Plots of the baseline hazard function, cumulative baseline hazard function, and baseline survival functions as well as their 95% confidence intervals are also available, as shown in Figure 1. The plot of the baseline hazard function in particular may be of interest to practioners. For instance, it may be of interest to see the increase in the risk of failure at the tail end of the follow-up period that is evident on this plot. Using the `predict.phmc_mpl` function, it is also possible to obtain fitted survival curves to compare the two treatment groups. Consider the case of a male with the centred sample median age of 0.5791. The comparative fitted survival curves for the two treatment groups are shown in Figure 2. Both plots indicate that the non-treatment group has a higher risk of failure than the treatment group at all points in time.

Figure 1: Estimates of (left-right) the baseline hazard function, cumulative baseline hazard function, and baseline survival function



Note that the grey lines indicate 95% confidence intervals, calculated using the delta method.

Figure 2: Comparative plot of estimates of the hazard function for each treatment group (top) and comparative plot of fitted survival curves for each treatment group.



6 Conclusion and Future Research

The aim of this thesis was to develop and evaluate a novel method for fitting a proportional hazards mixture cure model to partly-interval censored survival data. With contemporary advancements in medical treatments and technologies, the ability to account for the presence of a cured fraction makes the mixture cure model an important tool for the analysis of time-to-event data. However, fitting these models can be a complex process, because their likelihood function formulation requires some estimate of the non-parametric baseline survival or hazard function. Previous work in this area has largely been limited in scope to right-censored survival data, and has offered scarce options for obtaining a smooth estimate of the baseline hazard function (Sy & Taylor 2000, Peng & Dear 2000, Shen et al. 2019, Chen et al. 2019). Additionally, existing options for fitting the model in `R` are extremely limited. These drawbacks have restricted the practical and clinical utility of the mixture cure proportional hazards model for analysis of survival data. To address these shortcomings, this thesis drew on an existing body of work concerning the use of maximum penalised likelihood (MPL) estimation to fit proportional hazards regression models. This approach allowed for the straightforward inclusion of left and interval censored observations, and ensured a smooth estimate of the baseline hazard function could be obtained. Furthermore, the adoption of a Newton-MI algorithm for constrained optimisation satisfied the non-negativity constraint on the baseline hazard function without the risk of numerical instability or local maxima associated with other more ad-hoc methods used in the past. This thesis also presented asymptotic properties for the MPL estimates of the parameters, including the baseline hazard function estimate. Finally, a set of `R` functions was developed in order to easily fit the proposed model.

The results of the two simulation studies carried out to evaluate the proposed model demonstrate that MPL estimation of a proportional hazards mixture cure model produces satisfactory parameter estimates. The first simulation study compared the performance of the MPL method proposed in this thesis with the expectation-maximisation (EM) method of Peng & Dear (2000), which was available for implementation in `R` via the `smcure` package (Cai et al. 2012*b*). The results showed that the MPL method produced regression parameter estimates with comparable bias to the estimates produced by the existing EM method. The asymptotic variance estimates for the regression parameters produced using the MPL method performed as well as the bootstrapped variance estimated by `smcure`. The results of the study indicated that the MPL estimates of the cumulative baseline hazard function were, again, comparable if not better than the Breslow estimates provided by `smcure` in the majority of cases. Furthermore, the MPL method proposed here has the ability to provide inference on survival quantities such as the baseline hazard function, unlike `smcure`. The second simulation study investigated the performance of the model when fitted to partly-interval censored data. It found that the regression parameter estimates showed little bias when censoring was not extreme, especially the proportional hazards regression parameter. The estimates of the asymptotic variance were

likewise satisfactory, meaning that valid inference could be made on these parameters. Additionally, when censoring was not extreme and the sample size was reasonable, this model generally produced unbiased baseline hazard function estimates, and reasonable asymptotic standard error estimates for the baseline hazard.

Overall, this thesis marks a significant contribution to the existing body of research in the area of proportional hazards mixture cure models. It offers a new method for fitting this model to partly-interval censored survival data which also allows for a smooth estimate of the baseline hazard function to be obtained. The results of the simulations studies show that both regression parameter estimates and estimates of the baseline hazard function obtained via an MPL method are satisfactory, and that the method is comparable, if not preferable, to existing approaches such as Peng & Dear (2000). Additionally, the development of asymptotic properties for the parameter estimates allows for inference on both the regression parameters and the baseline hazard function or other survival quantities. This is a departure from many of the previously available methods, which commonly relied on bootstrapping to provide standard error estimates for regression parameters and largely disregarded inference on the baseline hazard function all together. Importantly, this thesis also contributes a set of `R` functions that can be used to fit the proposed model. As a result, the work done here could easily be made use of in a practical or clinical setting. Based on the review of the literature provided, it seems that the `R` implementation provided here would be the among the first to allow users to fit and make inferences on a proportional hazards mixture cure model using partly-interval censored data. More generally, this thesis highlights the utility of the MPL method in the context of survival analysis, especially for approaching complex problems like interval-censored data and data with cured fractions. It also demonstrates the use of the Newton-MI algorithm for constrained optimisation in the context of survival analysis, building on the work done in Ma et al. (2014) and Ma et al. (2019).

Despite the promising results presented in this thesis, there remains significant scope for further evaluation of this model and further research in this area. The simulation studies carried out here focused on evaluating the model in terms of the bias in the parameter estimates and the performance of the asymptotic estimates of the variance. However, these elements of the model could be more thoroughly evaluated in future simulation studies by considering a case with more than one covariate in the proportional hazards regression. Further simulation studies may also offer more insight into the systematic overestimation of the cured fraction that was observed in some simulation scenarios. Furthermore, future simulation studies would offer more comprehensive insight into the model performance were they to more carefully investigate the impact of knot selection and smoothing parameter selection on parameter estimation. Further investigation of the smoothing parameter selection specifically would be of value for two reasons. Firstly, the method for automatically selecting a smoothing parameter outlined in this thesis was not evaluated in terms of its impact on the performance of the estimators. Secondly, there may be some relationship between the smoothing parameter and issues with identifiability in

the model, as inappropriate choices of the smoothing parameter value may produce improper estimates of the conditional survival function (Corbiere et al. 2009). Research with a greater focus on evaluating the selection of the smoothing parameter for this model may therefore offer insight into how model performance could be improved.

Needless to say, the scope of this thesis did not include a number of areas that are currently of interest in the fields of cure model estimation, non-parametric baseline function estimation, and survival analysis as a whole. Future work on the model proposed in this thesis would be able to draw more broadly on this ongoing research. In doing so, it would enhance the utility and flexibility of the model. Firstly, an obvious current limitation of the method presented here is that it offers only one model for the incidence and one model for the latency. A next step might be to investigate MPL estimation of a mixture cure model that uses alternative models, or offers a range of models to choose from. Recent work by Li & Ma (2019) and Li & Ma (2020) explored MPL estimation of additive hazards models and accelerated failure time models. Expanding this into the context of a mixture cure model would widen the range of scenarios where the model in this thesis could be applied. Moreover, expansion of the choice of models available could bring this method more into line with the recent emphasis on semi-parametric transformation mixture cure models for partly-interval censored data, explored by Chen et al. (2019) and Shen et al. (2019). Future research might also be informed by recent developments in MPL estimation of proportional hazards models. One focus of recent research in this area has been on fitting the proportional hazards model in the presence of possibly dependent censoring. Work such as Xu et al. (2017) and Xu et al. (2018) has recently demonstrated the use of copulas in the MPL estimation of a proportional hazards regression model where the assumption of independent censoring is not satisfied. Another focus has been on the incorporation of time-varying covariates and competing risks (Thackham & Ma 2020*a*,*b*). Expansion the model presented in this thesis to incorporate dependent censoring, time-varying covariates or competing risks constitute promising future directions for research. A final potential avenue for future research may be to build on the presently small body of work focused on variable selection (Scolas et al. 2016) and diagnostic checks (Scolas et al. 2018, Muller & Van Keilegom 2019) for mixture cure models. Any of these possible future directions would add to the utility and flexibility of the model and estimation procedure presented in this thesis.

References

Amdahl, J. (2020), *flexsurvcure: Flexible Parametric Cure Models*. R package version 1.2.0.
URL: *https://CRAN.R-project.org/package=flexsurvcure*

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

Armijo, L. (1966), 'Minimisation of functions having Lipschitz continuous first partial derivatives', *Pacific Journal of Mathematics* **16**(1), 1–3.

Banerjee, S. & Carlin, B. P. (2004), 'Parametric spatial cure rate models for interval-censored time-to-relapse data', *Biometrics* **60**(1), 268–275.

Cai, C., Zou, Y., Peng, Y. & Zhang, J. (2012*a*), 'smcure: An R package for estimating semi-parametric mixture cure models', *Computer Methods and Programs in Biomedicine* **108**(3), 1255–1260.

Cai, C., Zou, Y., Peng, Y. & Zhang, J. (2012*b*), *smcure: Fit Semiparametric Mixture Cure Models*. R package version 2.0.
URL: *https://CRAN.R-project.org/package=smcure*

Cai, T. & Betensky, R. A. (2003), 'Hazard regression for interval-censored data with penalised spline', *Biometrics* **59**(3), 570–579.

Cantor, A. B. & Shuster, J. J. (1992), 'Parametric versus nonparametric methods for estimating cure rates based on censored survival data', *Statistics in Medicine* **11**(7), 931–937.

Chan, R. H. & Ma, J. (2012), 'A multiplicative-iterative algorithm for box-constrained penalised likelihood image restoration', *IEEE Transactions on Image Processing* **21**(7), 3168–3181.

Chen, C.-M., Shen, P.-s. & Huang, W.-L. (2019), 'Semiparametric transformation models for interval-censored data in the presence of a cure fraction', *Biometrical Journal* **61**(1), 203–215.

Corbiere, F., Commenges, D., Taylor, J. M. G. & Joly, P. (2009), 'A penalised likelihood approach for mixture cure models', *Statistics in Medicine* **28**, 510–524.

Couturier, D.-L., Ma, J., Heritier, S. & Manuguerra., M. (2017), *survivalMPL: Penalised Maximum Likelihood for Survival Analysis Models*. R package version 0.2.
URL: *https://CRAN.R-project.org/package=survivalMPL*

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

Cox, D. (1975), 'Partial likelihood', *Biometrika* **62**(2), 269–276.

Craven, P. & Wahba, G. (1979), 'Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalised cross-validation', *Numerische Mathematik* **31**, 377–402.

Farewell, V. (1982), 'The use of mixture models for the analysis of survival data with long-term survivors', *Biometrics* **38**(4), 1041–1046.

Farewell, V. (1986), 'Mixture models in survival analysis: Are they worth the risk?', *The Canadian Journal of Statistics* **14**(3), 257–262.

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

Gamel, J. & Voggel, R. (1997), 'Comparison of parametric and nonparametric survival methods using clinical data', *Statistics in Medicine* **16**(14), 1629–1643.

Ghittany, M., Maller, R. & Zhou, S. (1994), 'Exponential mixture models with long-term survivors and covariates', *Journal of Multivariate Analysis* **49**, 218–241.

Gordon, N. (1990), 'Application of the theory of finite mixtures for the estimation of 'cure' rates of treated cancer patients', *Statistics in Medicine* **9**(4), 397–407.

Gray, R. J. (1994), 'Spline-based tests in survival analysis', *Biometrics* **50**(3), 640–652.

Green, P. & Silverman, B. (1994), *Nonparametric Regression and Generalised Linear Models*, number 58 *in* 'Monographs on Statistics and Applied Probability', 1 edn, Chapman & Hall.

Huang, J. (1996), 'Efficient estimation for the proportional hazards model with interval censoring', *The Annals of Statistics* **24**(2), 540–568.

Jackson, C. (2016), 'flexsurv: A platform for parametric survival modelling in R', *Journal of Statistical Software* **70**(8), 1–33.

Joly, P., Commenges, D. & Letenneur, L. (1998), 'A penalised likelihood approach for arbitrarily censored and truncated data: Application to age-specific incidence of dementia', *Biometrics* **54**(1), 185–194.

Kauermann, G. (2005), 'A note on smoothing parameter selection for penalised spline smoothing', *Journal of Statistical Planning and Inference* **127**, 53–69.

Kauermann, G., Krivobokova, T. & Fahrmeir, L. (2009), 'Some asymptotic results on generalised penalised spline smoothing', *Journal of the Royal Statistical Society* **71**(2), 487–503.

Kimeldorf, G. S. & Wahba, G. (1970), 'A correspondence between Bayesian estimation on stochastic processes and smoothing by splines', *The Annals of Mathematical Statistics* **41**(2), 495–502.

Kirkwood, J. M., Strawderman, M. H., Ernstoff, M. S., Smith, T. J. & Blum, R. H. (1996), 'Interferon alfa-2b adjuvant therapy of high-risk respected cutaneous melanoma: the Eastern Cooperative Oncology Group Trial EST 1684', *Journal of Clinical Oncology* **14**(1), 7–17.

Krivobokova, T. (2013), 'Smoothing parameter selection in two frameworks for penalised splines', *Journal of the Royal Statistical Society* **75**(4), 725–741.

Kuk, A. Y. & Chen, C.-H. (1992), 'A mixture model combining logistic regression with proportional hazards regression', *Biometrika* **79**(3), 531–541.

Li, J. & Ma, J. (2019), 'Maximum penalised likelihood estimation of additive hazard models with partly interval censoring', *Computational Statistics & Data Analysis* **137**, 170–180.

Li, J. & Ma, J. (2020), 'On hazard-based penalised likelihood estimation of accelerated failure time model with partly interval censoring', *Statistical Methods in Medical Research* **29**(12), 3804–3817.

Ma, J. (2010), 'Positively constrained multiplicative iterative algorithm for maximum penalised likelihood tomographic reconstruction', *IEEE Transactions on Nuclear Science* **57**(1), 181–192.

Ma, J., Couturier, D.-L., Heritier, S. & Marschner, I. (2019), 'Proportional hazards model with partly interval censoring and its penalized likelihood estimation', *arXiv preprint arXiv:1904.06789* .

Ma, J., Heritier, S. & Lo, S. N. (2014), 'On the maximum penalised likelihood approach for proportional hazard models with right censored survival data', *Computational Statistics and Data Analysis* **74**, 142–156.

Maller, R. & Zhou, S. (1992), 'Estimating the proportion of immunes in a censored sample', *Biometrika* **79**(4), 731–739.

Muller, U. & Van Keilegom, I. (2019), 'Goodness-of-fit tests for the cure rate in a mixture cure model', *Biometrika* **106**(1), 211–227.

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

Patilea, V. & Van Keilegom, I. (2017), 'A general approach for cure models in survival analysis'.

Peng, Y. (2003), 'Estimating baseline distribution in proportional hazards cure models', *Computational Statistics and Data Analysis* **42**, 187–201.

Peng, Y. & Dear, K. B. G. (2000), 'A nonparametric mixture cure model for cure rate estimation', *Biometrics* **56**, 237–243.

Peng, Y., Dear, K. B. G. & Denham, J. (1998), 'Generalised F mixture model for cure rate estimation', *Statistics in Medicine* **17**, 813–830.

Ramsay, J. (1988), 'Monotone regression splines in action', *Statistical Science* **3**(4), 425–441.

Reiss, P. T. & Ogden, R. T. (2009), 'Smoothing parameter selection for a class of semiparametric linear models', *Journal of the Royal Statistical Society* **71**(2), 505–523.

Scolas, S., El Ghouch, A., Legrand, C. & Oulhaj, A. (2016), 'Variable selection in a flexible parametric mixture cure model with interval-censored data', *Statistics in Medicine* **35**, 1210–1225.

Scolas, S., Legrand, C., Oulhaj, A. & El Ghouch, A. (2018), 'Diagnostic checks in mixture cure models with interval censoring', *Statistical Methods in Medical Research* **27**(7), 2114–2132.

Shao, Q. & Zhou, X. (2004), 'A new parametric model for survival data with long-term survivors', *Statistics in Medicine* **23**(22), 3525–3543.

Shen, P.-s., Chen, H.-J., Pan, W.-H. & Chen, C.-M. (2019), 'Semiparametric regression analysis for left-truncated and interval-censored data without or with a cure fraction', *Computational Statistics and Data Analysis* **140**, 74–87.

Stein, M. L. (1990), 'A comparison of generalised cross validation and modified maximum likelihood for estimation the parameters of a stochastic process', *The Annals of Statistics* **18**(3), 1139–1157.

Sy, J. P. & Taylor, J. M. G. (2000), 'Estimation in a cox proportional hazards cure model', *Biometrics* **56**, 227–236.

Taylor, J. M. G. (1995), 'Semi-parametric estimation in failure time mixture models', *Biometrics* **51**(3), 899–907.

Thackham, M. & Ma, J. (2020*a*), 'On maximum likelihood estimation of competing risks using the cause-specific semi-parametric Cox model with time-varying covariates - an application of credit risk', *Journal of the Operational Research Society* .

Thackham, M. & Ma, J. (2020*b*), 'On maximum likelihood estimation of the semi-parametric Cox model with time-varying covariates', *Journal of Applied Statistics* **47**(9), 1511–1528.

Tsodikov, A., Ibrahim, J. G. & Yakovlev, A. Y. (2003), 'Estimating cure rates from survival data: An alternative to two-component mixture models', *Journal of the American Statistical Association* **98**(464), 1063–1078.

Wahba, G. (1978), 'Improper priors, spline smoothing and the problem of guarding against model errors in regression', *Journal of the Royal Statistical Society* **40**(3), 364–372.

Wahba, G. (1985), 'A comparison for GCV and GML for choosing the smoothing parameter in the generalised spline smoothing problem', *The Annals of Statistics* **13**(4), 1378–1402.

Wang, L., Du, P. & Liang, H. (2012), 'Two-component mixture cure rate model with spline estimated nonparametric components', *Biometrics* **68**, 726–735.

Wood, S. N. (2011), 'Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalised linear models', *Journal of the Royal Statistical Society* **73**(1), 3–36.

Xu, J., Ma, J., Connors, M. H. & Brodaty, H. (2018), 'Proportional hazard model estimation under dependent censoring using copulas and penalised likelihood', *Statistics in Medicine* **37**, 2238–2251.

Xu, J., Ma, J. & Prvan, T. (2017), 'Non-parametric hazard estimation with dependent censoring using penalised likelihood and an assumed copula', *Communications in Statistics - Theory and Methods* **46**(22), 11383–11403.

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

Yin, G. & Ibrahim, J. G. (2005), 'Cure rate models: A unified approach', *The Canadian Journal of Statistics* **33**(4), 559–570.

Yu, Y. & Ruppert, D. (2002), 'Penalised spline estimation for partially linear single-index models', *Journal of the American Statistical Association* **97**(460), 1042–1054.

Zeng, D., Yin, G. & Ibrahim, J. G. (2006), 'Semiparametric transformation models for survival data with a cure fraction', *Journal of the American Statistical Association* **101**(474), 670–684.

Zhou, J., Zhang, J., McLain, A. C. & Cai, B. (2016), 'A multiple imputation approach for semiparametric cure model with interval censored data', *Computational Statistics and Data Analysis* **99**, 105–114.

A Elements of the score vector and Hessian matrix

The components of the score vector are as follows. Let z_{it} be element t of vector $\mathbf{z}_i$ and x_{ij} be element j of vector $\mathbf{x}_i$, for $t = 1, ..., p$ and $j = 1, ..., q$. Then, the first derivative of $\Phi(\boldsymbol{\eta})$ with respect to β_t is

$$\frac{\partial \Phi(\boldsymbol{\eta})}{\partial \beta_t} = \sum_{i=1}^{n} z_{it} \left((1-\delta_i^R)(1-\pi(\mathbf{z}_i)) + \delta_i^R \left(\frac{\pi(\mathbf{z}_i)(1-\pi(\mathbf{z}_i))(S(t_i|\mathbf{x}_i)-1)}{1-\pi(\mathbf{z}_i)+\pi(\mathbf{z}_i)S(t_i|\mathbf{x}_i)} \right) \right)$$

The first derivative of $\Phi(\boldsymbol{\eta})$ with respect to γ_j is

$$\begin{aligned}\frac{\partial \Phi(\boldsymbol{\eta})}{\partial \gamma_j} = \sum_{i=1}^{n} x_{ij} \Bigg(& \delta_i(1-H(t_i|\mathbf{x}_i)) - \delta_i^R \left(\frac{\pi(\mathbf{z}_i)S(t_i|\mathbf{x}_i)H(t_i|\mathbf{x}_i)}{1-\pi(\mathbf{z}_i)-\pi(\mathbf{z}_i)S(t_i|\mathbf{x}_i)} \right) \\ & + \delta_i^L \left(\frac{S(t_i|\mathbf{x}_i)H(t_i|\mathbf{x}_i)}{1-S(t_i|\mathbf{x}_i)} \right) - \delta_i^I \left(\frac{S(t_i^L|\mathbf{x}_i)H(t_i^L|\mathbf{x}_i) - S(t_i^R|\mathbf{x}_i)H(t_i^R|\mathbf{x}_i)}{S(t_i^L|\mathbf{x}_i) - S(t_i^R|\mathbf{x}_i)} \right) \Bigg)\end{aligned}$$

The first derivative of $\Phi(\boldsymbol{\eta})$ with respect to θ_w, where $w = 1, ..., m$, is

$$\begin{aligned}\frac{\partial \Phi(\boldsymbol{\eta})}{\partial \theta_w} = \sum_{i=1}^{n} \Bigg(& \delta_i \left(\frac{\psi_w(t_i)}{h_0(t)} - e^{\mathbf{x}_i^T\boldsymbol{\gamma}}\Psi_w(t_i) \right) \\ & - \delta_i^R \left(\frac{\pi(\mathbf{z}_i)S(t_i|\mathbf{x}_i)\Psi_w(t_i)e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}{1-\pi(\mathbf{z}_i)+\pi(\mathbf{z}_i)S(t_i|\mathbf{x}_i)} \right) + \delta_i^L \left(\frac{S(t_i|\mathbf{x}_i)\Psi_w(t_i)e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}{1-S(t_i|\mathbf{x}_i)} \right) \\ & - \delta_i^I \left(\frac{S(t_i^L|\mathbf{x}_i)\Psi_w(t_i^L)e^{\mathbf{x}_i^T\boldsymbol{\gamma}} - S(t_i^R|\mathbf{x}_i)\Psi_w(t_i^R)e^{\mathbf{x}_i^T\boldsymbol{\gamma}}}{S(t_i^L|\mathbf{x}_i) - S(t_i^R|\mathbf{x_i})} \right) \Bigg) - \lambda \frac{\partial J(\boldsymbol{\eta})}{\partial \theta_w}\end{aligned}$$

where

$$\lambda \frac{\partial J(\boldsymbol{\eta})}{\partial \theta_w} = 2\lambda \boldsymbol{\theta} \mathbf{R}$$

when using the penalty function discussed above.

The components of the Hessian matrix are as follows. The second derivative of $\Phi(\boldsymbol{\eta})$ with respect to β_u and β_t is

$$\begin{aligned}\frac{\partial^2 \Phi(\boldsymbol{\eta})}{\partial \beta_u \partial \beta_t} = & -\sum_{i=1}^{n} z_{it} z_{iu} ((1-\delta_i^R)\pi(\mathbf{z}_i)(1-\pi(\mathbf{z}_i)) \\ & - \delta_i^R (S(t_i|\mathbf{x}_i)-1)\pi(\mathbf{z}_i)(1-\pi(\mathbf{z}_i)) \left(\frac{(1-\pi(\mathbf{z}_i))^2 - S(t_i|\mathbf{x}_i)\pi(\mathbf{z}_i)^2}{(1-\pi(\mathbf{z}_i)+\pi(\mathbf{z}_i)S(t_i|\mathbf{x}_i))^2} \right)\end{aligned}$$

The second derivative of $\Phi(\boldsymbol{\eta})$ with respect to β_t and γ_j is

$$\frac{\partial^2 \Phi(\boldsymbol{\eta})}{\partial \gamma_j \partial \beta_t} = -\sum_{i=1}^{n} x_{ij} z_{it} \delta_i^R \frac{H(t_i|\mathbf{x}_i)S(t_i|\mathbf{x}_i)\pi(\mathbf{z}_i)(1-\pi(\mathbf{z}_i))}{(1-\pi(\mathbf{z}_i)+S(t_i|\mathbf{x}_i)\pi(\mathbf{z}_i))^2}$$

The second derivative of $\Phi(\boldsymbol{\eta})$ with respect to β_t and θ_w is

$$\frac{\partial^2 \Phi(\boldsymbol{\eta})}{\partial \theta_w \partial \beta_t} = -\sum_{i=1}^{n} z_{it} e^{\mathbf{x}_i^T\boldsymbol{\gamma}} \delta_i^R \frac{S(t_i|\mathbf{x}_i)\Psi_w(t_i)\pi(\mathbf{z}_i)(1-\pi(\mathbf{z}_i))}{(1-\pi(\mathbf{z}_i)+S(t_i|\mathbf{x}_i)\pi(\mathbf{z}_i))^2}$$

The second derivative of $\Phi(\boldsymbol{\eta})$ with respect to γ_j and γ_k is

$$\begin{aligned}\frac{\partial^2\Phi(\boldsymbol{\eta})}{\partial\gamma_j\partial\gamma_k} &= -\sum_{i=1}^{n} x_{ij}x_{ik}(\delta_i H(t_i|\mathbf{x}_i) + \delta_i^R \frac{\pi(\mathbf{z}_i)S(t_i|\mathbf{x}_i)H(t_i|\mathbf{x}_i)(1-H(t_i|\mathbf{x}_i))}{1-\pi(\mathbf{z}_i)+S(t_i|\mathbf{x}_i)\pi(\mathbf{z}_i)} \\ &+ \delta_i^R \frac{\pi(\mathbf{z}_i)^2 S(t_i|\mathbf{x}_i)^2 H(t_i|\mathbf{x}_i)^2}{(1-\pi(\mathbf{z}_i)+S(t_i|\mathbf{x}_i)^2} + \delta_i^L \frac{S(t_i|\mathbf{x}_i)H(t_i|\mathbf{x}_i)(S(t_i|\mathbf{x}_i)+H(t_i|\mathbf{x_i})-1)}{(1-S(t_i|\mathbf{x}_i))^2} \\ &+ \delta_i^I \frac{S(t_i^L|\mathbf{x}_i)H(t_i^L|\mathbf{x}_i)(1-H(t_i^L|\mathbf{x}_i)) - S(t_i^R|\mathbf{x}_i)H(t_i^R|\mathbf{x}_i)(1-H(t_i^R|\mathbf{x}_i))}{S(t_i^L|\mathbf{x}_i)-S(t_i^R|\mathbf{x}_i)} \\ &+ \delta_i^I \frac{S(t_i^L|\mathbf{x}_i)^2 H(t_i^L|\mathbf{x}_i)^2 - S(t_i^R|\mathbf{x}_i)^2 H(t_i^R|\mathbf{x}_i)^2}{(S(t_i^L|\mathbf{x}_i)-S(t_i^R|\mathbf{x}_i))^2}\end{aligned}$$

The second derivative of $\Phi(\boldsymbol{\eta})$ with respect to γ_j and θ_w is

$$\begin{aligned}\frac{\partial^2\Phi(\boldsymbol{\eta}}{\partial\gamma_j\partial\theta_w} &= -\sum_{i=1}^{n} x_{ij}e^{\mathbf{x}_i^T\gamma}(\delta_i\Psi_w(t_i) + \delta_i^R\Psi_w(t_i)S(t_i|\mathbf{x}_i)\frac{\pi(\mathbf{z}_i)S(t_i|\mathbf{x}_i)+(1-\pi(\mathbf{z}_i))(1-H(t_i|\mathbf{x}_i))}{(1-\pi(\mathbf{z}_i)+S(t_i|\mathbf{x}_i)\pi(\mathbf{z}_i))^2} \\ &+ \delta_i^L S(t_i|\mathbf{x}_i)\Psi_w(t_i)\left(\frac{H(t_i|\mathbf{x}_i)+S(t_i|\mathbf{x}_i)-1}{(1-S(t_i|\mathbf{x}_i))^2}\right) \\ &+ \delta_i^I \frac{S(t_i^L|\mathbf{x}_i)\Psi(t_i^L)(1-H(t_i^L|\mathbf{x}_i)) - S(t_i^R|\mathbf{x}_i)\Psi(t_i^R)(1-H(t_i^R|\mathbf{x}_i))}{S(t_i^L|\mathbf{x}_i)-S(t_i^R|\mathbf{x}_i)} \\ &+ \delta_i^I \frac{\left(S(t_i^L|\mathbf{x}_i)\Psi_w(t_i^L) - S(t_i^R|\mathbf{x}_i)\Psi_w(t_i^R)\right)\left(S(t_i^L|\mathbf{x}_i)H(t_i^L|\mathbf{x}_i) - S(t_i^R|\mathbf{x}_i)H(t_i^R|\mathbf{x}_i)\right)}{(S(t_i^L|\mathbf{x}_i)-S(t_i^R|\mathbf{x}_i))^2}\end{aligned}$$

Finally, the second derivative of $\Phi(\boldsymbol{\eta})$ with respect to θ_w and θ_z is

$$\begin{aligned}\frac{\partial^2\Phi}{\partial\theta_w\partial\theta_z} &= -\sum_{i=1}^{n}\delta_i\frac{\psi_w(t_i)\psi_z(t_i)}{h_0^2(t_i)} - e^{2X_i^T\gamma}\Bigg(\delta_i^R\frac{\pi(Z_i)(1-\pi(Z_i))\Psi_w(t_i)\Psi_z(t_i)S(t_i)}{(1-\pi(Z_i)-S(t_i)\pi(Z_i))^2} \\ &- \delta_i^L\frac{\Psi_w(t_i)\Psi_z(t_i)S(t_i)}{(1-S(t_i))^2} - \delta_i^I S(t_i^L)S(t_i^R)\frac{[\Psi_w(t_i^R)-\Psi_w(t_i^L)][\Psi_z(t_i^R)-\Psi_z(t_i^L)]}{(S(t_i^L)-S(t_i^R))^2}\Bigg)\end{aligned}$$

B Simulation Study 1 Supplementary Tables

Table B1: Number of knots selected for each simulation scenario

Sample size n	Censoring distribution parameter λ_c	Approximate $\pi(\mathbf{z})$	Number of knots
100	4.2	$0.5, 0.75$	3, 5
	1.25	$0.5, 0.75$	4, 4
500	4.2	$0.5, 0.75$	6, 8
	1.25	$0.5, 0.75$	5, 8
2000	4.2	$0.5, 0.75$	7, 10
	1.25	$0.5, 0.75$	5, 6

Table B2: 95% coverage probabilities of β_1 and γ_1

n		100	100	500	500	2000	2000
$\pi(\mathbf{z})$		0.5	0.75	0.5	0.75	0.5	0.75
π^R		0.59	0.41	0.59	0.41	0.59	0.41
$\hat{\beta}_1$	MPL	0.96	0.96	0.94	0.96	0.89	0.95
	EM	0.97	1.00	0.92	0.96	0.89	0.95
$\hat{\gamma}_1$	MPL	0.93	0.96	0.99	0.92	0.93	0.96
	EM	0.95	0.97	0.97	0.92	0.92	0.95
π^R		0.75	0.64	0.75	0.64	0.75	0.64
$\hat{\beta}_1$	MPL	0.93	0.95	0.94	0.98	0.93	0.93
	EM	0.94	0.97	0.96	0.94	0.93	0.94
$\hat{\gamma}_1$	MPL	0.91	0.92	0.98	0.92	0.96	0.98
	EM	0.98	0.96	0.97	0.98	0.96	0.98

Note that the values of π^R presented here are the proportions of right censored observations in the whole sample i.e. including the right censored observations from individuals in the cured fraction. Scenarios with π^R of 0.59 or 0.41 correspond to scenarios with the censoring distribution parameter $\lambda_c = 4.2$. Scenarios with π^R of 0.75 or 0.64 correspond to scenarios with $\lambda_c = 1.25$.

C Simulation Study 2 Supplementary Tables

Table C1: Number of knots selected for each simulation scenario

Sample size n	Non-cured fraction event probability π^E	Approximate $\pi(\mathbf{z})$	Number of knots
100	0.5	0.5, 0.75	4, 6
	0.25	0.5, 0.75	4, 5
500	0.5	0.5, 0.75	5, 8
	0.25	0.5, 0.75	4, 7
2000	0.5	0.5, 0.75	5, 8
	0.25	0.5, 0.75	4, 6

Table C2: 95% coverage probabilities of β_0, β_1 and γ_1

n	100	100	500	500	2000	2000
$\pi(\mathbf{z})$	0.5	0.75	0.5	0.75	0.5	0.75
π^E	0.25	0.36	0.25	0.36	0.25	0.36
π^R	0.67	0.52	0.67	0.52	0.67	0.52
$\hat{\beta}_0$	0.95	0.98	0.96	0.97	0.96	0.95
$\hat{\beta}_1$	0.93	0.97	0.93	0.94	0.98	0.94
$\hat{\gamma}_1$	0.96	0.94	0.98	0.95	0.96	0.98
π^E	0.13	0.18	0.13	0.18	0.13	0.18
π^R	0.75	0.64	0.75	0.64	0.75	0.64
$\hat{\beta}_0$	0.71	0.77	0.23	0.20	0.00	0.00
$\hat{\beta}_1$	0.96	0.96	0.96	0.96	0.88	0.86
$\hat{\gamma}_1$	1.00	0.99	0.96	0.98	0.97	0.98

Note that the values of π^E and π^R presented here are the proportions of events and right censored observations in the whole sample i.e. including right censored observations from individuals in the cured fraction. Scenarios with a π^E of 0.25 or 0.36 and a π^R of 0.67 or 0.52 correspond to an event probability in the non-cured fraction of 50%. Scenarios with a π^E of 0.13 or 0.18 and a π^R of 0.75 or 0.64 correspond to an event probability in the non-cured fraction of 25%.

Table C3: 95% coverage probabilities of the estimate of $h_0(t)$

n	100	100	500	500	2000	2000
$\pi(\mathbf{z})$	0.5	0.75	0.5	0.75	0.5	0.75
π^E	0.25	0.36	0.25	0.36	0.25	0.36
π^R	0.67	0.52	0.67	0.52	0.67	0.52
t_1	0.93	0.95	0.93	0.98	0.92	0.96
t_2	0.98	0.95	0.96	0.98	0.97	0.94
t_3	0.97	0.97	0.97	0.99	0.95	0.95
π^E	0.13	0.18	0.13	0.18	0.13	0.18
π^R	0.75	0.64	0.75	0.64	0.75	0.64
t_1	0.97	1.00	0.84	0.92	0.17	0.42
t_2	0.97	0.94	0.80	0.96	0.24	0.49
t_3	0.97	0.96	0.98	0.99	0.74	0.84

D Access to `R` package

Access to the `R` package developed as part of this thesis, discussed in Chapter 5, is available at the following link: `https://github.com/annabelwebb/thesis_submission`