

A Non-Mixture Cure Model with a Change Point in a Co-Variate for Right Censored Data

Durga H. Kotal^{1*}, Lianfen Qian²

¹CSM Department of Mathematics, Augusta University, Augusta, USA

²Department of Mathematics, The University of Arizona, Tucson, USA

Email: *dudkotal@augusta.edu, lqian01@arizona.edu

How to cite this paper: Kotal, D.H. and Qian, L.F. (2026) A Non-Mixture Cure Model with a Change Point in a Co-Variate for Right Censored Data. *Open Journal of Statistics*, 16, 16-37.
<https://doi.org/10.4236/ojs.2026.161002>

Received: January 14, 2026

Accepted: February 6, 2026

Published: February 9, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

We study a non-mixture cure model with a covariate change-point for right-censored survival data and develop maximum-likelihood estimation under a smoothed likelihood to handle the non-differentiability induced by the threshold. Assuming exponential latency for susceptibles, we derive closed-form scores through a stable reparameterization and jointly estimate the change-point, cure fraction, and rates. In simulations spanning multiple sample sizes, censoring levels, and covariate distributions, the estimator exhibits small bias and competitive RMSE, with accurate change-point recovery. To assess robustness, we further conduct sensitivity analyses under latency misspecification, in which susceptible failure times follow a Weibull distribution while the model is fitted assuming exponential latency; the results show that estimation of the cure fractions and change-point remains stable, whereas the latency rate parameters converge to pseudo-true values as expected under misspecification. We illustrate the method on two biomedical datasets: (i) the colon cancer dataset using the number of positive lymph nodes as the threshold covariate, and (ii) a melanoma cohort using Breslow thickness. In both applications, the fitted model provides clinically interpretable strata, with subgroup-specific cure fractions that are negligible to modest and a threshold estimate consistent with established prognostic cut-offs. These results demonstrate that the smoothed change-point non-mixture cure model is practical, interpretable, and reliable for detecting threshold effects in the presence of long-term survivors.

Keywords

Non-Mixture Model, Change Point, Maximum Likelihood Method, Smoothed Likelihood, Right Censored Survival Data, Covariate

1. Introduction

The cure fraction models are broadly used for analyzing survival data where a significant portion of the population may be cured and thus not subject to the event of interest. In the existing literature, there are two major approaches to model survival data with cure fraction. The first is mixture cure rate model, introduced by Boag [1] in 1949 and further developed by Berkson and Gage [2] in 1952 and later studied extensively by several authors. The second type is the non-mixture cure rate model, also referred to as the bounded cumulative hazard model or the promotion time cure model. In oncology research, the latter model was developed under the assumption that the number of cancer cells remaining active after treatment, which can grow slowly and eventually lead to detectable recurrence, follows a Poisson distribution. This assumption was first proposed by Yakovlev [3] in 1993 and was further discussed by Chen [4] in 1999, Ibrahim and Chen [5] in 2002 and Tsodikov [6] in 2002. Tsodikov [7] in 2003 provided a review of existing methodology of statistical inference based on the non-mixture model. Kutal and Qian [8] in 2018 studied a non-mixture cure model for right censored data with Fréchet distribution.

Change-point models for detection and estimation have been widely studied in the field of survival analysis for decades. Matthews and Farewell [9] in 1982 first introduced the change point in constant hazard rate when analyzing the failure times of non-lyphoblastic leukemia patients. Muller and Wang [10] in 1990 proposed a non-parametric approach for the change in hazard rates for censored survival data. They used the kernel method for the non-parametric estimation of derivative of hazard rate. Pons [11] in 2003 considered a Cox regression model with a change-point according to a threshold in a co-variate. Duppy [12] in 2006 studied estimation in a change-point hazard regression model with co-variate. Li and Qian [13] in 2013 estimated a change point hazard regression model with long term survivors. Zhao [14] in 2009 analyzed the change-point model for survival data including long term survivors. Othus [15] in 2012 studied the change-point cure models based on co-variate threshold. Taweab [16] in 2015 investigated the Bounded Cumulative Hazard (BCH) model with a change-point at an unknown threshold in a co-variate for right censored data. Kutal [17] in 2018 examined parameter estimation for a non-mixture cure model incorporating a change-point in a covariate under right-censored data. The current study builds upon and extends that earlier work by refining the estimation framework and broadening its application. However, in the literature, there is not much work in the field of cure models with change-point and only few studies have investigated this type of model. In this article, we propose an approach to analyze the non-mixture cure model in the presence of right censored data with a change point according to a threshold in a covariate and we considered exponential distribution for susceptible subjects.

2. Non-Mixture Cure Model and Susceptible Distribution

2.1. Model Motivation and Formulation

We introduced a non-mixture cure model developed by Yakovlev [3] in 1993 as an

alternative to the mixture cure model. The motivation for this model is as follows. Let N denote the number of potential malignant cancer cells that may eventually grow and produce a detectable tumor at a time in the future. We assumed N follow a Poisson distribution with mean $\lambda > 0$. Let $Z_j, j = 1, 2, \dots, N$ represent the random time until that the j th cancer cell to produce a detectable cancer mass. We assume the times $Z_j, j = 1, 2, \dots, N$ are independent and identically distributed random variables. The time to relapse in cancer for an individual is defined as $T = \min\{Z_j, j = 1, 2, \dots, N\}$. Let an individual cumulative distribution function (CDF) $F(t|x)$ and a survival function $S(t|x) = 1 - F(t|x)$, where x is a relevant scalar covariate. The population survival function of the $S_T(t|x)$, which represents the probability of no cancer relapse by time t , is derived as follows:

$$\begin{aligned} S_T(t|x) &= P[\text{No cancer by time } t|x] \\ &= P[N = 0|x] + P[Z_1 > t, Z_2 > t, \dots, Z_N > t, N \geq 1|x] \\ &= e^{-\lambda(x)} + \sum_{N=1}^{\infty} S^N(t|x) \frac{\lambda^N(x)}{N!} e^{-\lambda(x)} \\ &= e^{-\lambda(x) + \lambda(x)S(t|x)} = e^{-\lambda(x)F(t|x)} \end{aligned} \tag{1}$$

The cure fraction, $p(x)$, is the probability of long-term survival (having no malignant cells ($N = 0$)), which can be defined as

$$p(x) = \lim_{t \rightarrow \infty} S_T(t|x) = P(N = 0|x) = e^{-\lambda(x)}.$$

Notice that as $\lambda(x) \rightarrow 0$, we obtain $p(x) \rightarrow 1$ where as $\lambda(x) \rightarrow \infty$, we obtain $p(x) \rightarrow 0$. Since $\lim_{t \rightarrow \infty} S_T(t|x) = e^{-\lambda(x)} \geq 0$, then the model (1) is an improper survival function. Moreover, the survival function of T also can be written as $S_T(t|x) = p(x)^{F(t|x)}$. From this, the population's cumulative distribution function, probability density function, and hazard function are $F_T(t|x) = 1 - p(x)^{F(t|x)}$, $f_T(t|x) = -\ln p(x) f(t|x) p(x)^{F(t|x)}$, and $h_T(t|x) = -\ln p(x) f(t|x)$ respectively.

For right censored survival time, the observed time for the i^{th} individual is $y_i = \min(T_i, C_i)$, where C_i is the censoring time. A censoring indicator, δ_i , is defined as 1 if the event (failure) was observed and 0 if the data was right-censored.

The likelihood function for a sample of n individual is given by

$$L = \prod_{i=1}^n h_T^{\delta_i}(y_i|x_i) S_T(y_i|x_i) = \prod_{i=1}^n [-\ln p(x_i) f(y_i|x_i)]^{\delta_i} p(x_i)^{F(y_i|x_i)} \tag{2}$$

The corresponding log-likelihood function is

$$l = \ln(L) = \sum_{i=1}^n \delta_i \ln[-\ln p(x_i)] + \sum_{i=1}^n \delta_i \ln f(y_i|x_i) + \sum_{i=1}^n \ln p(x_i) F(y_i|x_i) \tag{3}$$

2.2. Exponential Distribution for Uncured Individuals

In this paper, we assume that the failure times of uncured individuals follow an exponential distribution, a common choice for lifetime data analysis, with rate parameter ($\mu > 0$). The exponential distribution of the probability density function, the cumulative distribution function, the survival function, and the hazard

function are $f(y|x) = \mu e^{-\mu y}$; $y \geq 0$, $F(y|x) = 1 - e^{-\mu y}$; $y \geq 0$, $S(y|x) = e^{-\mu y}$; $y \geq 0$, and $h(y|x) = \mu$; $y \geq 0$, respectively.

For simulation study, to generate random numbers from a given distribution, we apply the Inverse Transform Method.

Let $Y \sim U(0,1)$ and $F(x), x \in \mathbb{R}$, be a cumulative distribution function, then $F^{-1}(y) = \min\{x: F(x) \geq y\}$, $y \in [0,1]$. The random variable $X = F^{-1}(Y)$ is a continuous random variable with cumulative distribution function $F(x)$. Since $p(X \leq x) = p(F^{-1}(Y) \leq x) = p(Y \leq F(x)) = F(x)$, since Y is a uniform distribution.

3. The Log-Likelihood of Non-Mixture Model with a Change Point in a Covariate

We extend the non-mixture cure model to incorporate a change point, denoted by τ , in a continuous covariate x . The model divide the population into two distinct regimes: (i) individuals with $x_i \leq \tau$, the cure probability p_1 , the failure rate μ_1 and $S(y|x) = S_1(y|x)$ (ii) individuals with $x_i > \tau$, the cure probability p_2 , the failure rate μ_2 and $S(y|x) = S_2(y|x)$. The likelihood contribution of the i^{th} individual is

$$L_i = \begin{cases} \left[-\ln(p_1) f_1(y_i | x_i) \right]^{\delta_i} p_1^{F_1(y_i | x_i)}, & \text{if } x_i \leq \tau, \\ \left[-\ln(p_2) f_2(y_i | x_i) \right]^{\delta_i} p_2^{F_2(y_i | x_i)}, & \text{if } x_i > \tau. \end{cases}$$

The complete observed data are (y_i, δ_i, x_i) and the unknown parameters are defined by $\theta^T = (p_1, p_2, \mu_1, \mu_2, \tau)$. The likelihood function under change point τ can be written as

$$L(\theta) = \prod_{i=1}^n \left\{ \left[-\ln(p_1) f_1(y_i | x_i) \right]^{\delta_i} p_1^{F_1(y_i | x_i)} \right\}^{I(x_i \leq \tau)} * \left\{ \left[-\ln(p_2) f_2(y_i | x_i) \right]^{\delta_i} p_2^{F_2(y_i | x_i)} \right\}^{I(x_i > \tau)} \quad (4)$$

Taking the logarithm, the corresponding log-likelihood function can be written as

$$l(\theta) = \sum_{i=1}^n \left\{ I(x_i \leq \tau) \left[\delta_i \ln(-\ln(p_1)) + \delta_i \ln f_1(y_i | x_i) + F_1(y_i | x_i) \ln(p_1) \right] + [1 - I(x_i \leq \tau)] \left[\delta_i \ln(-\ln(p_2)) + \delta_i \ln f_2(y_i | x_i) + F_2(y_i | x_i) \ln(p_2) \right] \right\} \quad (5)$$

where, $f_1(y_i | x_i) = \mu_1 e^{-\mu_1 y_i}$, $\ln(f_1(y_i | x_i)) = \ln(\mu_1) - \mu_1 y_i$, and $\ln(f_2(y_i | x_i)) = \ln(\mu_2) - \mu_2 y_i$. The log likelihood function (5) is not differentiable with respect to the change point parameter τ . This lack of smoothness complicates the use of standard gradient-based optimization methods, motivating the introduction of a smoothed likelihood approach, which is discussed in the next section.

3.1. Smoothed Likelihood Method

To address the non-differentiability of the log-likelihood function in the presence

of a change point, we employ a smoothed likelihood approach. This method replaces the discontinuous indicator functions $I(x_i \leq \tau)$ and $I(x_i > \tau)$ with a continuous and differentiable approximation by Othus *et al.* [15] in 2012. For this, we use a continuous and differential function $k(\cdot)$ satisfying

$\lim_{u \rightarrow -\infty} k(u) = 0$ and $\lim_{u \rightarrow \infty} k(u) = 1$. For a given sample size n , define the smoothed version $k_n(u) = k\left(\frac{u}{h_n}\right)$, where $h_n > 0$ is a bandwidth parameter that

controls the degree of smoothing and may depend on n . In our simulation studies, we set $h_n = n^{-\gamma}$, $\gamma > 0$, where γ is chosen so that it balances finite sample performance and ensures asymptotic properties of the estimator. For the asymptotic results, we assume the bandwidth satisfies $h_n \rightarrow 0$ and $\sqrt{nh_n} \rightarrow \infty$ as $n \rightarrow \infty$. This condition ensures that the smoothed objective is asymptotically equivalent to the original change-point likelihood while retaining differentiability, and it yields root- n consistency and asymptotic normality for the full parameter vector, including τ . For sufficiently small h_n , we have $k_n(x_i - \tau) \rightarrow 1$ when $(x_i - \tau) > 0$ and $k_n(x_i - \tau) \rightarrow 0$ when $(x_i - \tau) < 0$. Thus, the smooth model preserves the essential structure of the change point while providing differentiability. A commonly used choice for this class of function, $k(\cdot)$, is the logistic function, defined as

$$k_n(u) = \frac{\exp\left(\frac{u}{h_n}\right)}{1 + \exp\left(\frac{u}{h_n}\right)}$$

Using this formulation, the smoothed likelihood function associated for the observed data (y_i, δ_i, x_i) is given by

$$L^*(\theta) = \prod_{i=1}^n \left\{ \left[-\ln(p_1) f_1(y_i | x_i) \right]^{\delta_i} p_1^{F_1(y_i | x_i)} \right\}^{k_n(\tau - x_i)} * \left\{ \left[-\ln(p_2) f_2(y_i | x_i) \right]^{\delta_i} p_2^{F_2(y_i | x_i)} \right\}^{1 - k_n(\tau - x_i)} \quad (6)$$

Taking logarithms, the smoothed log-likelihood function becomes

$$l^*(\theta) = \sum_{i=1}^n \left\{ k_n(\tau - x_i) \left[\delta_i \ln(-\ln(p_1)) + \delta_i \ln f_1(y_i | x_i) + F_1(y_i | x_i) \ln(p_1) \right] + \left[1 - k_n(\tau - x_i) \right] \left[\delta_i \ln(-\ln(p_2)) + \delta_i \ln f_2(y_i | x_i) + F_2(y_i | x_i) \ln(p_2) \right] \right\} \quad (7)$$

We proposed to estimate the parameters, including change point τ , by maximizing the smoothed log-likelihood Equation (7) using Newton-Raphson algorithm. The smoothing parameter h_n plays critical role in this procedure. A smaller choice of h_n leads to estimates that are nearly unbiased but may suffer from higher variability, whereas larger values of h_n reduce the variance at the expense of increased bias. Thus, an appropriate balance of h_n is required to achieve stable and reliable estimation. The detailed estimation framework, including re-parameterization, score function, observed information, and algorithmic considerations, is presented in the Estimation Procedure section to provide a sys-

tematic approach for implementation.

3.2. Estimation Procedure

The parameters $\theta = (p_1, p_2, \mu_1, \mu_2, \tau)$ of the proposed non-mixture cure model with a change point are estimated by maximizing smoothed log-likelihood function $\ell^*(\theta)$ introduced earlier section. The smoothing process ensures that the function is differentiable with respect to the parameters, including the change point τ , making standard optimization feasible. The complete estimation process is detailed below.

Re-parameterization for stability: To perform unconstrained optimization and ensure numerical stability, the model's parameters are re-parameterized. The original parameter vector $\theta = (p_1, p_2, \mu_1, \mu_2, \tau)$ is subject to the constraints $p_j \in (0,1)$ and $\mu_j > 0$. The cure probabilities p_j are transformed using the logit function

$$a_j = \text{logit}(p_j), \quad p_j = \frac{\exp(a_j)}{1 + \exp(a_j)}, \quad j = 1, 2,$$

and the failure rate parameters μ_j are transformed using the log function

$$b_j = \ln(\mu_j), \quad \mu_j = \exp(b_j), \quad j = 1, 2.$$

This creates an unconstrained working parameter vector $\eta = (a_1, a_2, b_1, b_2, \tau)^\top$, where each element can take any value in the real line. The original parameters can be recovered via the inverse transformations.

Score function (first derivatives): The smoothed log-likelihood is

$$\ell^*(\theta) = \sum_{i=1}^n [k_i t_{1i} + (1 - k_i) t_{2i}], \quad k_i = k_n(\tau - x_i),$$

where t_{1i} and t_{2i} denote the log-likelihood contributions for individuals with $x_i \leq \tau$ and $x_i > \tau$, respectively. The score vector, $g(\eta)$, is the gradient of the smoothed log-likelihood with respect to the unconstrained parameter vector η . It is defined as $g(\eta) = \nabla_{\eta} \ell^*(\eta)$ and the components of the score vector are:

$$\begin{aligned} \frac{\partial \ell^*}{\partial a_1} &= \sum_{i=1}^n k_i \frac{\partial t_{1i}}{\partial a_1}, & \frac{\partial \ell^*}{\partial a_2} &= \sum_{i=1}^n (1 - k_i) \frac{\partial t_{2i}}{\partial a_2}, \\ \frac{\partial \ell^*}{\partial b_1} &= \sum_{i=1}^n k_i \frac{\partial t_{1i}}{\partial b_1}, & \frac{\partial \ell^*}{\partial b_2} &= \sum_{i=1}^n (1 - k_i) \frac{\partial t_{2i}}{\partial b_2}, \\ \frac{\partial \ell^*}{\partial \tau} &= \frac{1}{h_n} \sum_{i=1}^n k_i (1 - k_i) (t_{1i} - t_{2i}). \end{aligned}$$

Observed information (second derivatives): The observed information matrix, $H(\eta)$, is defined as the negative of the Hessian matrix of the smoothed log-likelihood:

$$H(\eta) = -\nabla_{\eta}^2 \ell^*(\eta),$$

where $\ell^*(\eta) = \sum_{i=1}^n \{k_n(\tau - x_i) t_{1i} + [1 - k_n(\tau - x_i)] t_{2i}\}$ with t_{ji} is the log-likelihood contribution for the i^{th} subject in regime $j \in \{1, 2\}$:

$$t_{ji} = \delta_i \ln \left\{ -\ln(p_j) \right\} + \delta_i (\ln \mu_j - \mu_j y_i) + \ln(p_j) F_j(y_i)$$

This matrix plays an essential role in the Newton-Raphson estimation procedure and is used to obtain the asymptotic variance and standard errors of the parameter estimates. As a result of the reparameterization $\eta = (a_1, a_2, b_1, b_2, \tau)^\top$, the Hessian naturally partitions into blocks. The second derivative with respect to (a_1, b_1) and (a_2, b_2) are available in closed form analytic expressions because the corresponding terms in the smoothed log-likelihood depend directly on the exponential density and the logit link, whose derivatives are tractable. In contrast, the derivatives involving the change-point parameter τ includes the smoothing function $k_i = k_n(\tau - x_i)$, whose first and second derivatives, k_i' and k_i'' , enters into the Hessian. These cross-derivative terms are algebraically more complex. Therefore, in practice, it is common to evaluate the τ related Hessian elements numerically for enhanced numerical stability, while retaining analytic forms for the remaining blocks. A complete set of analytic derivatives, including the score vector and Hessian components, is provided in the Supplement for reproducibility and implementation.

Newton-Raphson algorithm: The parameter vector η that maximizes the smoothed log-likelihood function $\ell^*(\eta)$ is found using the Newton-Raphson algorithm. Starting from an initial value $\eta^{(0)}$, the estimate is refined through successive iterations. For the t -th iteration, the standard update is calculated as:

$$\eta^{(t+1)} = \eta^{(t)} - \left[H(\eta^{(t)}) \right]^{-1} g(\eta^{(t)}).$$

To ensure robust convergence and prevent overshooting the maximum, a step-size parameter $\alpha \in (0, 1]$ is often introduced, modifying the update to:

$$\eta^{(t+1)} = \eta^{(t)} - \alpha \left[H(\eta^{(t)}) \right]^{-1} g(\eta^{(t)}).$$

This iterative process continues until convergence is achieved, which is determined when the norm of the score vector $g(\eta)$ and the magnitude of the change between successive parameter estimates both fall below a pre-specified tolerance.

Because the Newton-Raphson algorithm can be sensitive to starting values, we employed a multi-start initialization strategy. Initial values for the change point τ were selected from a coarse grid over empirical quantiles of the threshold covariate, and for each candidate τ the remaining parameters were initialized using simple regime-specific summaries, including empirical late-time survival proportions for the cure components and exponential rate estimates for the failure components. To mitigate convergence to local maxima, the algorithm was run from multiple starting values, and the solution yielding the largest smoothed log-likelihood was retained. Convergence was assessed using both the norm of the score vector and the change in parameter estimates, with step-size damping applied as needed to stabilize the updates.

Asymptotic Properties: The estimator $\hat{\theta}$ is defined as the maximizer of the smoothed log-likelihood

$$\ell_n^*(\theta) = \sum_{i=1}^n \ell_i^*(\theta)$$

in (7), where

$$\theta = (p_1, p_2, \mu_1, \mu_2, \tau)^\top.$$

Although our model is a non-mixture (promotion time) cure model with parametric exponential latency, and thus differs in formulation from the change-point cure models in Othus *et al.* (2012) and the BCH change-point model in Taweab *et al.* (2015), the asymptotic argument follows the same smoothed M-estimation framework: consistency is obtained from uniform convergence and identifiability, and asymptotic normality follows from a Taylor expansion of the smoothed score and a central limit theorem for the i.i.d. score contributions.

Assume standard regularity conditions for M-estimators hold (e.g., identifiability, interior true parameter, finite second moments, and nonsingularity of the information matrix), and that the smoothing bandwidth satisfies

$$h_n \rightarrow 0 \text{ and } \sqrt{nh_n} \rightarrow \infty.$$

Then $\hat{\theta}$ is consistent and

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{H}(\theta_0)^{-1}),$$

where θ_0 is the true parameter and

$$\mathcal{H}(\theta_0) = -\mathbb{E}[\nabla_{\theta}^2 \ell_n^*(\theta_0)]$$

is the expected (smoothed) Fisher information. In practice, $\mathcal{H}(\theta_0)$ is consistently estimated by the observed information

$$-\nabla_{\theta}^2 \ell_n^*(\hat{\theta}),$$

yielding the usual large-sample covariance approximation

$$\widehat{\text{Var}}(\hat{\theta}) = \{-\nabla_{\theta}^2 \ell_n^*(\hat{\theta})\}^{-1}.$$

Because the optimization is performed on the unconstrained scale

$$\eta = (a_1, a_2, b_1, b_2, \tau)^\top,$$

with $a_j = \text{logit}(p_j)$ and $b_j = \log(\mu_j)$, standard errors for

$$(p_1, p_2, \mu_1, \mu_2, \tau)$$

are obtained by applying the delta method to the inverse observed-information matrix on the η -scale.

4. Simulation Study

In this section, we conducted a simulation study to evaluate the finite-sample performance of the proposed non-mixture cure model with a change point in covariate based on right censored data. The sample data were generated from a non-mixture cure model with a change point according to the following procedure:

Step-1: We considered two distinct scenarios for the covariate distribution and the true change point τ . (i) The covariate x_i was drawn from a uniform distribution $U(0, 2)$ with a true change point at $\tau = 1$. (ii) The covariate x_i was drawn

from a truncated distribution on the interval $(0, 3)$ with a mean of 1 and a standard deviation of 1, with a true change point at $\tau = 1.5$.

Step-2: For each simulated subject, a random variable u was drawn from the uniform distribution $U(0, 1)$. If $u \leq p(x_i)$, the subject is assigned to the cure group, otherwise the subject is assigned to the uncured group.

Step 3: (Inverse-CDF draw of T_i)

Draw $U_i \sim \text{Uniform}(0,1)$. We use the population survival $S_T(t) = \exp\{-\lambda F(t)\}$ with exponential latency $F(t) = 1 - e^{-\mu t}$. Set $U_i = S_T(t)$ and solve:

$$U_i = \exp\{-\lambda F(t)\} \Rightarrow F(t) = -\frac{1}{\lambda} \ln U_i \Rightarrow 1 - e^{-\mu t} = -\frac{1}{\lambda} \ln U_i.$$

This has a finite solution for t only when $U_i \in (e^{-\lambda}, 1)$, which correspond to the susceptible population, yielding

$$t_i = -\frac{1}{\mu} \ln \left(1 + \frac{1}{\lambda} \ln U_i \right), \quad U_i \in (e^{-\lambda}, 1). \quad (8)$$

If $U_i \leq e^{-\lambda}$, classify the subject as cured and set $T_i = \infty$.

For the Change-point implementation, with regimes $j \in \{1, 2\}$ by $x_i \leq \tau$ (regime 1) or $x_i > \tau$ (regime 2), use the regime specific parameters $(\lambda, \mu) = (\lambda_j, \mu_j)$ in the above formula.

Step-4: Apply independent right-censoring by generating censoring times C_i from an exponential distribution with rate parameter $\mu > 0$, where μ is chosen to achieve a desired censoring level.

Step-5: Estimate the parameter vector $\hat{\theta}$ by maximizing the smoothed log-likelihood function using standard optimization procedures.

For each scenario, we simulated 1000 replications with sample sizes of 100, 200, and 400. The performance of the estimators was assessed using the mean, bias, standard error (SE), and root mean square error (RMSE). The simulation results demonstrate that the proposed estimation method performs well across all settings. The average parameter estimates are close to their true values, with small biases. Estimation of the parameter of the change point τ is generally accurate, even under moderate censoring. As expected, SE and RMSE decrease as the sample size increases from 100 to 200 and 200 to 400, confirming the consistency of the estimators. In addition, the method performs better under lower levels of censoring, with reduced variability and bias compared to higher censoring rates. Overall, the results indicate that the proposed smoothed likelihood approach provides reliable parameter estimates in finite samples.

The detailed numerical summaries of the simulation results are reported in **Table 1** and **Table 2**. **Table 1** presents the finite-sample performance of the proposed estimator under a Uniform $U(0, 2)$ covariate distribution with a true change point at $\tau = 1$, while **Table 2** summarizes results under a truncated normal covariate distribution with $\tau = 1.5$. In both scenarios, the estimators exhibit small bias and decreasing standard errors and RMSE as the sample size increases, with improved performance under lower censoring rates.

Table 1. Simulation summaries under $U(0, 2)$ with $\tau = 1$ at two censoring levels.

| (a) Censoring $\approx 27\%$, $\tau = 1$ | | | | | | (b) Censoring $\approx 41\%$, $\tau = 1$ | | | | | |
|---|------------|--------|--------|--------|--------|---|------------|--------|--------|--------|--------|
| Parameter | θ_0 | Mean | Bias | SE | RMSE | Parameter | θ_0 | Mean | Bias | SE | RMSE |
| $N = 100$ | | | | | | $N = 100$ | | | | | |
| p_1 | 0.02 | 0.0215 | 0.0015 | 0.0128 | 0.0129 | p_1 | 0.02 | 0.0221 | 0.0021 | 0.0145 | 0.0147 |
| μ_1 | 0.30 | 0.3041 | 0.0041 | 0.0952 | 0.0953 | μ_1 | 0.30 | 0.3069 | 0.0069 | 0.1158 | 0.1160 |
| p_2 | 0.01 | 0.0109 | 0.0009 | 0.0076 | 0.0077 | p_2 | 0.01 | 0.0113 | 0.0013 | 0.0088 | 0.0089 |
| μ_2 | 0.40 | 0.4077 | 0.0077 | 0.1103 | 0.1106 | μ_2 | 0.40 | 0.4102 | 0.0102 | 0.1341 | 0.1345 |
| τ | 1.00 | 1.0189 | 0.0189 | 0.1854 | 0.1864 | τ | 1.00 | 1.0253 | 0.0253 | 0.2011 | 0.2027 |
| $N = 200$ | | | | | | $N = 200$ | | | | | |
| p_1 | 0.02 | 0.0208 | 0.0008 | 0.0089 | 0.0089 | p_1 | 0.02 | 0.0211 | 0.0011 | 0.0099 | 0.0100 |
| μ_1 | 0.30 | 0.3015 | 0.0015 | 0.0651 | 0.0651 | μ_1 | 0.30 | 0.3031 | 0.0031 | 0.0789 | 0.0789 |
| p_2 | 0.01 | 0.0104 | 0.0004 | 0.0051 | 0.0051 | p_2 | 0.01 | 0.0106 | 0.0006 | 0.0060 | 0.0060 |
| μ_2 | 0.40 | 0.4036 | 0.0036 | 0.0765 | 0.0766 | μ_2 | 0.40 | 0.4049 | 0.0049 | 0.0922 | 0.0923 |
| τ | 1.00 | 1.0095 | 0.0095 | 0.1288 | 0.1292 | τ | 1.00 | 1.0136 | 0.0136 | 0.1452 | 0.1458 |
| $N = 400$ | | | | | | $N = 400$ | | | | | |
| p_1 | 0.02 | 0.0203 | 0.0003 | 0.0061 | 0.0061 | p_1 | 0.02 | 0.0205 | 0.0005 | 0.0068 | 0.0068 |
| μ_1 | 0.30 | 0.3009 | 0.0009 | 0.0453 | 0.0453 | μ_1 | 0.30 | 0.3017 | 0.0017 | 0.0546 | 0.0546 |
| p_2 | 0.01 | 0.0102 | 0.0002 | 0.0035 | 0.0035 | p_2 | 0.01 | 0.0103 | 0.0003 | 0.0041 | 0.0041 |
| μ_2 | 0.40 | 0.4018 | 0.0018 | 0.0532 | 0.0532 | μ_2 | 0.40 | 0.4025 | 0.0025 | 0.0645 | 0.0646 |
| τ | 1.00 | 1.0041 | 0.0041 | 0.0901 | 0.0902 | τ | 1.00 | 1.0067 | 0.0067 | 0.1018 | 0.1020 |

Note: θ_0 is true parameter value; SE = standard error; RMSE = root mean square error.

Table 2. Simulation summaries under Truncated Normal(1, 1) on $[0, 3]$ with $\tau = 1.5$ at two censoring levels.

| (a) Censoring $\approx 27\%$, $\tau = 1.5$ | | | | | | (b) Censoring $\approx 42\%$, $\tau = 1.5$ | | | | | |
|---|------------|--------|--------|--------|--------|---|------------|--------|--------|--------|--------|
| Parameter | θ_0 | Mean | Bias | SE | RMSE | Parameter | θ_0 | Mean | Bias | SE | RMSE |
| $N = 100$ | | | | | | $N = 100$ | | | | | |
| p_1 | 0.02 | 0.0218 | 0.0018 | 0.0139 | 0.0140 | p_1 | 0.02 | 0.0225 | 0.0025 | 0.0158 | 0.0160 |
| μ_1 | 0.30 | 0.3051 | 0.0051 | 0.1021 | 0.1022 | μ_1 | 0.30 | 0.3081 | 0.0081 | 0.1255 | 0.1258 |
| p_2 | 0.01 | 0.0110 | 0.0010 | 0.0081 | 0.0082 | p_2 | 0.01 | 0.0115 | 0.0015 | 0.0096 | 0.0097 |
| μ_2 | 0.40 | 0.4088 | 0.0088 | 0.1197 | 0.1200 | μ_2 | 0.40 | 0.4119 | 0.0119 | 0.1480 | 0.1485 |
| τ | 1.50 | 1.5201 | 0.0201 | 0.2235 | 0.2244 | τ | 1.50 | 1.5287 | 0.0287 | 0.2451 | 0.2468 |
| $N = 200$ | | | | | | $N = 200$ | | | | | |
| p_1 | 0.02 | 0.0209 | 0.0009 | 0.0095 | 0.0095 | p_1 | 0.02 | 0.0213 | 0.0013 | 0.0108 | 0.0109 |
| μ_1 | 0.30 | 0.3023 | 0.0023 | 0.0703 | 0.0703 | μ_1 | 0.30 | 0.3040 | 0.0040 | 0.0863 | 0.0864 |
| p_2 | 0.01 | 0.0105 | 0.0005 | 0.0055 | 0.0055 | p_2 | 0.01 | 0.0107 | 0.0007 | 0.0066 | 0.0066 |
| μ_2 | 0.40 | 0.4041 | 0.0041 | 0.0831 | 0.0832 | μ_2 | 0.40 | 0.4059 | 0.0059 | 0.1011 | 0.1013 |
| τ | 1.50 | 1.5108 | 0.0108 | 0.1569 | 0.1573 | τ | 1.50 | 1.5151 | 0.0151 | 0.1705 | 0.1712 |

Continued

| | $N = 400$ | | | | | $N = 400$ | | | | | |
|---------|-----------|--------|--------|--------|--------|-----------|------|--------|--------|--------|--------|
| p_1 | 0.02 | 0.0204 | 0.0004 | 0.0065 | 0.0065 | p_1 | 0.02 | 0.0206 | 0.0006 | 0.0075 | 0.0075 |
| μ_1 | 0.30 | 0.3011 | 0.0011 | 0.0490 | 0.0490 | μ_1 | 0.30 | 0.3021 | 0.0021 | 0.0602 | 0.0602 |
| p_2 | 0.01 | 0.0102 | 0.0002 | 0.0038 | 0.0038 | p_2 | 0.01 | 0.0103 | 0.0003 | 0.0046 | 0.0046 |
| μ_2 | 0.40 | 0.4020 | 0.0020 | 0.0583 | 0.0583 | μ_2 | 0.40 | 0.4029 | 0.0029 | 0.0708 | 0.0709 |
| τ | 1.50 | 1.5049 | 0.0049 | 0.1095 | 0.1096 | τ | 1.50 | 1.5078 | 0.0078 | 0.1199 | 0.1202 |

Note: θ_0 is true parameter value; SE = standard error; RMSE = root mean square error.

Figure 1 provides a graphical summary of the simulation results by illustrating the behavior of the RMSE as a function of sample size for all model parameters across the considered scenarios. As shown in **Figure 1**, the RMSE decreases monotonically as the sample size increases for each parameter, confirming the consistency of the proposed smoothed maximum likelihood estimators under both covariate distributions and censoring levels.

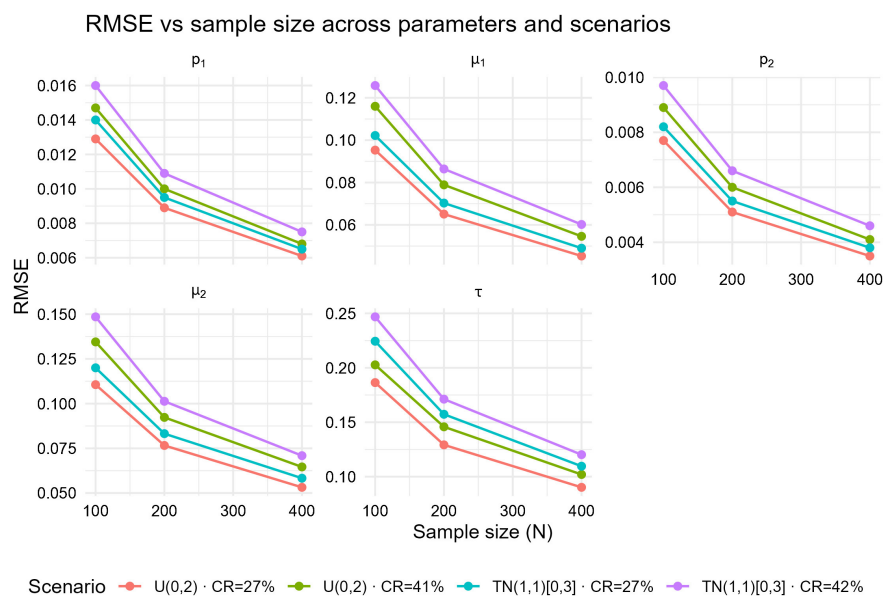


Figure 1. RMSE decreases with increases sample sizes for all parameters.

Sensitivity and Misspecification Analysis

A crucial test of the model’s reliability is its robustness when the core assumption of an exponential latency distribution for susceptible individuals is violated. We performed a comprehensive sensitivity analysis where the true survival times were generated using a Weibull distribution (with shape parameter $\alpha = 2.0$, implying an increasing hazard) while the estimation was carried out using the exponential-based smoothed maximum likelihood estimator (SMLE). To further examine robustness with respect to the covariate design, this analysis was conducted under both the Uniform ($U(0, 2)$) and Truncated Normal ($TN(1, 1)$ on $[0,3]$) covariate

distributions to ensure generalizability.

As summarized in **Table 3**, the fitted model consistently exhibits significant structural bias in the failure rate estimates ($\hat{\mu}_1, \hat{\mu}_2$), with typical biases ranging from approximately -0.17 to -0.27 across all scenarios. This outcome is expected, as the constant-hazard exponential model cannot accurately capture the time-varying hazard of the true Weibull process; consequently, the rate estimator converges to pseudo-true values under misspecification. Crucially, the key parameters of clinical interest of the cure fractions (\hat{p}_1, \hat{p}_2) remain robust and approximately unbiased across all sample sizes and covariate distributions. Most notably, the change-point estimator $\hat{\tau}$ demonstrates strong robustness, with negligible bias and steadily decreasing RMSE as N grows (for example, from 0.2409 to 0.1136 in the $U(0, 2)$ scenario). These results indicate that, even under substantial latency misspecification, the proposed smoothed change-point non-mixture cure model reliably identifies threshold effects and provides robust inference for parameters of primary scientific interest.

Table 3. Sensitivity analysis under latency misspecification. Susceptible failure times are generated from a Weibull distribution, while the model is fitted assuming exponential latency.

| (a) Uniform(0, 2), $\tau_0 = 1$ | | | | | | (b) TruncNorm[0, 3], $\tau_0 = 1.5$ | | | | | |
|---------------------------------|------------|--------|---------|--------|--------|-------------------------------------|------------|--------|---------|--------|--------|
| Parameter | θ_0 | Mean | Bias | SE | RMSE | Parameter | θ_0 | Mean | Bias | SE | RMSE |
| N = 100 | | | | | | N = 100 | | | | | |
| p_1 | 0.02 | 0.0161 | -0.0039 | 0.0180 | 0.0185 | p_1 | 0.02 | 0.0125 | -0.0075 | 0.0151 | 0.0168 |
| p_2 | 0.01 | 0.0080 | -0.0020 | 0.0100 | 0.0102 | p_2 | 0.01 | 0.0101 | 0.0001 | 0.0142 | 0.0142 |
| μ_1 | 0.30 | 0.1265 | -0.1735 | 0.0427 | 0.1787 | μ_1 | 0.30 | 0.1127 | -0.1873 | 0.0390 | 0.1913 |
| μ_2 | 0.40 | 0.1344 | -0.2656 | 0.0419 | 0.2688 | μ_2 | 0.40 | 0.1378 | -0.2622 | 0.0455 | 0.2661 |
| τ | 1.00 | 1.0130 | 0.0130 | 0.2406 | 0.2409 | τ | 1.50 | 1.5213 | 0.0213 | 0.2948 | 0.2956 |
| N = 200 | | | | | | N = 200 | | | | | |
| p_1 | 0.02 | 0.0077 | -0.0123 | 0.0098 | 0.0157 | p_1 | 0.02 | 0.0083 | -0.0117 | 0.0106 | 0.0158 |
| p_2 | 0.01 | 0.0049 | -0.0051 | 0.0059 | 0.0078 | p_2 | 0.01 | 0.0057 | -0.0043 | 0.0071 | 0.0083 |
| μ_1 | 0.30 | 0.1059 | -0.1941 | 0.0302 | 0.1965 | μ_1 | 0.30 | 0.1034 | -0.1966 | 0.0304 | 0.1990 |
| μ_2 | 0.40 | 0.1252 | -0.2748 | 0.0326 | 0.2767 | μ_2 | 0.40 | 0.1270 | -0.2730 | 0.0342 | 0.2751 |
| τ | 1.00 | 1.0083 | 0.0083 | 0.1679 | 0.1681 | τ | 1.50 | 1.5033 | 0.0033 | 0.2418 | 0.2418 |
| N = 400 | | | | | | N = 400 | | | | | |
| p_1 | 0.02 | 0.0072 | -0.0128 | 0.0079 | 0.0150 | p_1 | 0.02 | 0.0066 | -0.0134 | 0.0084 | 0.0158 |
| p_2 | 0.01 | 0.0033 | -0.0067 | 0.0039 | 0.0078 | p_2 | 0.01 | 0.0047 | -0.0053 | 0.0058 | 0.0078 |
| μ_1 | 0.30 | 0.1064 | -0.1936 | 0.0254 | 0.1953 | μ_1 | 0.30 | 0.1006 | -0.1994 | 0.0260 | 0.2011 |
| μ_2 | 0.40 | 0.1213 | -0.2787 | 0.0256 | 0.2799 | μ_2 | 0.40 | 0.1275 | -0.2725 | 0.0333 | 0.2745 |
| τ | 1.00 | 1.0052 | 0.0052 | 0.1134 | 0.1136 | τ | 1.50 | 1.5376 | 0.0376 | 0.1466 | 0.1513 |

Note. θ_0 denotes the reference parameter value; SE = standard error; RMSE = root mean square error.

5. Real Data Application

5.1. The Colon Cancer Dataset

To illustrate the proposed methodology, we analyzed the colon cancer dataset from the survival package. The dataset records recurrence-free survival times, censoring indicators, and multiple clinical covariates. After basic cleaning (finite positive times, finite node counts), the analysis set contained $n = 911$ individuals with 456 recurrences and 455 censored observations. In our analysis, the number of positive lymph nodes (nodes) served as the threshold covariate for the change-point model.

5.1.1. Stratified Descriptive Analysis

Patients were further stratified into two groups according to the estimated change point. **Table 4** presents the descriptive summaries. The low-node group ($x \leq \hat{\tau}$) had a median time of 2023 days with an event rate of 41.3%, compared to 501.5 days and 66.5% in the high-node group.

Table 4. Descriptive summary of colon cancer patients by change-point stratification.

| Group | n | Events | Event Rate (%) | Median Time (days) |
|-------------------------|-----|--------|----------------|--------------------|
| $\leq \hat{\tau}$ nodes | 595 | 246 | 41.3 | 2023.0 |
| $> \hat{\tau}$ nodes | 316 | 210 | 66.5 | 501.5 |

5.1.2. Parameter Estimates

Table 5 reports the estimated cure fractions (p_1, p_2), failure rates (μ_1, μ_2), and the change point τ . The estimated change point was located at approximately $\hat{\tau} = 3.82$ positive nodes, partitioning the cohort into two prognostic subgroups. Patients with $x_i \leq \hat{\tau}$ were governed by parameters (p_1, μ_1) , while those with $x_i > \hat{\tau}$ followed (p_2, μ_2) . Standard errors (SE) were obtained via the observed information matrix using the delta method.

Table 5. Parameter estimates from the non-mixture cure model with change point (colon dataset).

| Parameter | Estimate | SE | Interpretation |
|--------------|----------|----------|---|
| p_1 | 0.542340 | 0.024631 | Cure fraction, low-node group ($x \leq \hat{\tau}$) |
| p_2 | 0.304869 | 0.034681 | Cure fraction, high-node group ($x > \hat{\tau}$) |
| μ_1 | 0.000996 | 0.000112 | Failure rate, low-node group |
| μ_2 | 0.001444 | 0.000157 | Failure rate, high-node group |
| $\hat{\tau}$ | 3.816639 | 0.378346 | Estimated change point (nodes) |

5.1.3. Profile Likelihood of the Change Point

Figure 2 displays the profile negative log-likelihood as a function of τ . The curve shows a clear minimum at $\hat{\tau} \approx 3.82$, providing evidence for a change point in the number of positive nodes.

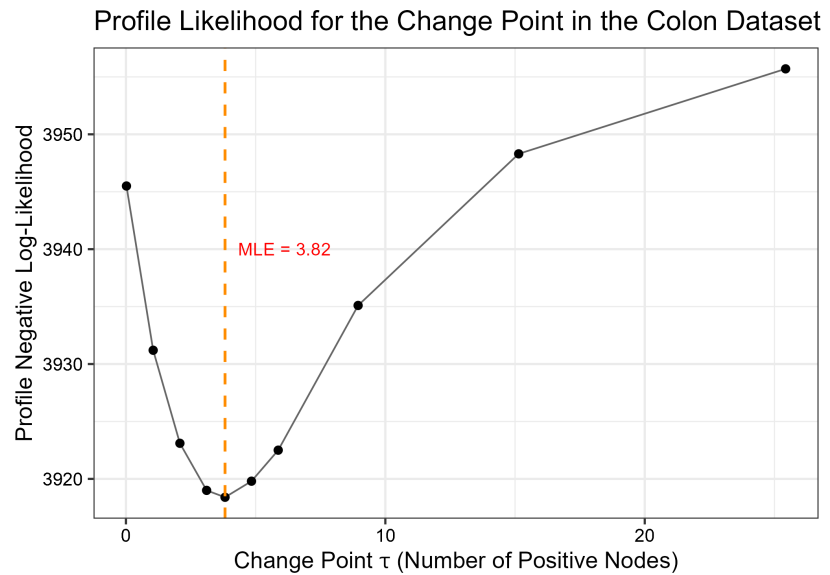


Figure 2. Profile negative log-likelihood for τ (colon dataset).

5.1.4. Interpretation

These findings confirm the critical prognostic role of lymph node involvement in colon cancer. The estimated cure fraction is substantially higher ($p_1 \approx 0.54$) and the susceptible failure rate is lower ($\mu_1 \approx 9.96 \times 10^{-4} \text{ day}^{-1}$) than above the change point ($p_2 \approx 0.30$, $\mu_2 \approx 1.44 \times 10^{-3} \text{ day}^{-1}$). The profile likelihood supports a well-defined change point near 3.8 nodes, and the stratified descriptive statistics reinforced the model-based conclusions. Overall, the analysis demonstrates the practical utility of the proposed smoothed non-mixture cure model with a change point for quantifying clinically interpretable thresholds in prognosis.

5.2. The Melanoma Dataset

To further validate the proposed methodology, we analyzed the melanoma dataset from the boot package ($n = 205$). The dataset records survival time in days, event status (1 = death due to melanoma, 0 = censored), and several clinical covariates including tumor thickness, ulceration, and age. In this application, tumor thickness (mm) was used as the change-point covariate.

5.2.1. Stratified Descriptive Analysis

Patients were stratified by the estimated change point. **Table 6** provides summary statistics. Those with thin lesions ($\leq \hat{\tau}$) showed higher cure probability, lower event rates, and longer median survival compared to the thick-lesion group.

Table 6. Descriptive summary of melanoma patients by change-point stratification.

| Group | n | Events | Event Rate (%) | Median Time (days) |
|--------------------------------|-----|--------|----------------|--------------------|
| Thin lesions (≤ 2.26 mm) | 118 | 18 | 15.3 | 2103.5 |
| Thick lesions (> 2.26 mm) | 87 | 39 | 44.8 | 1812.0 |

5.2.2. Parameter Estimates

Patients were stratified using the estimated change point $\hat{\tau} \approx 2.26$ mm. As summarized in **Table 6**, the thin-lesion group ($x \leq \hat{\tau}$) exhibited a substantially lower event rate (15.3%) than the thick-lesion group (44.8%) and a longer median survival time (2103.5 vs. 1812.0 days). Although these descriptive measures indicate a more favorable prognosis for patients with thin lesions, the model-based results in Section 5.2.1 suggest that this advantage is driven primarily by a markedly lower failure rate rather than a large cure fraction, implying substantially slower failure dynamics in this subgroup. **Table 7** reports the estimated cure fractions (p_1, p_2), failure rates (μ_1, μ_2), and the change point τ . The estimated change point was $\hat{\tau} \approx 2.26$ mm. Patients with tumors of thickness $\leq \hat{\tau}$ were described by parameters (p_1, μ_1) , whereas those with thicker tumors followed (p_2, μ_2) . Standard errors (SE) were computed using the observed information matrix with the delta method.

Table 7. Parameter estimates from the non-mixture cure model with change point (Melanoma dataset).

| Parameter | Estimate | SE | Interpretation |
|-----------|----------|---------|---|
| p_1 | 0.000045 | 0.0020 | Cure fraction, thin lesions (≤ 2.26 mm) |
| p_2 | 0.2965 | 0.1852 | Cure fraction, thick lesions (> 2.26 mm) |
| μ_1 | 0.000006 | 0.00003 | Failure rate, thin lesions |
| μ_2 | 0.000255 | 0.00018 | Failure rate, thick lesions |
| τ | 2.26 | 0.50 | Estimated change point (mm) |

5.2.3. Profile Likelihood of the Change Point

Figure 3 shows the profile negative log-likelihood for τ . The curve attains a clear minimum at $\hat{\tau} \approx 2.26$ mm, supporting the presence of a clinically meaningful threshold in tumor thickness.

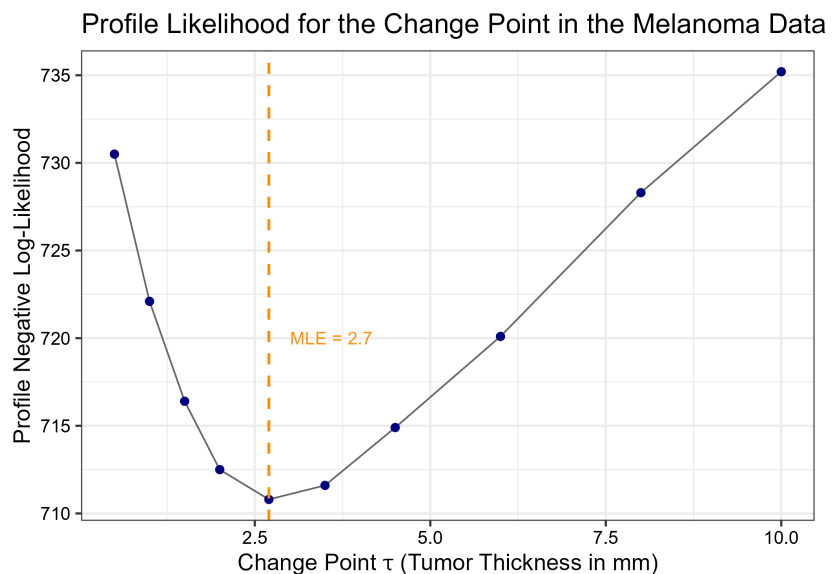


Figure 3. Profile negative log-likelihood for τ (Melanoma dataset).

5.2.4. Interpretation

These results confirm tumor thickness as a significant prognostic threshold in melanoma. While clinical intuition often associates thin lesions with a higher likelihood of being cured, the model-based estimates for the thin lesion group ($x \leq \hat{\tau} \approx 2.26$ mm) yield a cure fraction p_1 near zero. This indicates that the superior survival outcomes observed in this group characterized by a substantially lower event rate of 15.3% and a notably longer median survival time of 2103.5 days are statistically better captured by an exceptionally low failure rate ($\mu_1 = 0.000006$) rather than a definitive cured plateau. In contrast, patients with thick lesions ($x > \hat{\tau} \approx 2.26$ mm) experienced a higher failure rate ($\mu_2 = 0.000255$), yet the model identifies a more pronounced cure fraction ($p_2 = 0.2965$) for this subgroup. The proposed smoothed likelihood non-mixture cure model successfully identifies this threshold effect, demonstrating that when a distinct plateau is not yet visible in the data, the model relies on the latency distribution to characterize long-term survivors through slow failure dynamics. This highlights the necessity of interpreting cure fractions as model based constructs whose values are sensitive to follow-up patterns and parametric assumptions.

Interpretation of the change-point cure model

A critical consideration in the application of cure models is the reliable separation of long-term survivors from individuals with very long but finite survival times. In the melanoma application, the estimated cure fraction should therefore be interpreted as a model-based quantity rather than as definitive biological evidence of cure, particularly in the absence of a visible survival plateau or when follow-up is limited. Under such conditions, cure fraction estimates can be sensitive to the assumed parametric form of the latency distribution. For subgroups in which the estimated cure fraction is close to zero, the observed survival advantage is more plausibly explained by markedly slow failure dynamics among susceptible individuals rather than by the presence of a clearly identifiable cured subpopulation. Consequently, cure fraction estimates should be evaluated jointly with the estimated failure rates and the available follow-up duration to avoid overinterpretation of the cure component.

Limitation of univariate threshold modeling

While the proposed model successfully identifies clinically interpretable thresholds in both the colon cancer and melanoma datasets, the real-data analyses are conducted in a univariate setting. In clinical practice, survival outcomes are typically influenced by multiple prognostic factors; for example, melanoma prognosis depends not only on tumor thickness but also on ulceration status, patient age, and anatomical site. Excluding such covariates may introduce residual confounding, particularly when they are correlated with the threshold covariate, which may in turn affect the estimated change point $\hat{\tau}$. Moreover, the regime-specific cure fractions and failure rates (p_j, μ_j) should be interpreted as marginal effects of the threshold covariate, as they may reflect the combined influence of unmodeled prognostic factors. Extending the proposed smoothed likelihood framework to a multivariable setting to accommodate additional covariates within regimes repre-

sents an important direction for future research.

While the proposed model effectively identifies threshold effects, it assumes that covariates influence survival only through regime membership defined by the change point τ . In complex clinical settings, additional prognostic effects may persist within these regimes, and ignoring such effects may limit model flexibility. Acknowledging this limitation, future research should extend the proposed smoothed likelihood framework to accommodate within-regime covariate effects, potentially through semiparametric or partially parametric modeling approaches.

6. Conclusion

In this article, we developed and validated a non-mixture cure model with a covariate dependent change point for right censored survival data. Assuming exponential distributions for susceptible subjects, we employed a smoothed likelihood approach to address the non-smoothness in the likelihood due to the change point in covariate. The performance and reliability of our proposed estimation method were first confirmed through extensive simulation studies. The results demonstrated that the estimators for the cure fractions, failure rates, and the change-point parameter are consistent and exhibit low bias across a variety of realistic scenarios, including different sample sizes and censoring proportions. Additional sensitivity analyses under latency misspecification indicate that inference on the change-point and cure fractions remains robust even when the exponential assumption for susceptible survival times is violated. In both real-data applications, the profile likelihood analysis provided strong visual evidence for a unique and well-defined change point, validating the stability of the estimates. Application to the colon and melanoma datasets demonstrated the model's clinical utility by identifying significant thresholds in lymph node count and tumor thickness that sharply distinguish patient prognosis. This approach provides a valuable tool for refining risk stratification in clinical settings where such threshold effects are common.

Acknowledgements

The authors acknowledge the use of AI tools (Gemini, developed by Google) to support language editing and refinement during the preparation of this manuscript. All content and interpretations are the responsibility of the authors. The authors sincerely thank the anonymous reviewers for their careful reading of the manuscript and for their constructive comments and insightful suggestions, which have substantially improved the quality and clarity of this work.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Boag, J.W. (1949) Maximum Likelihood Estimates of the Proportion of Patients

- Cured by Cancer Therapy. *Journal of the Royal Statistical Society Series B. Statistical Methodology*, **11**, 15-44. <https://doi.org/10.1111/j.2517-6161.1949.tb00020.x>
- [2] Berkson, J. and Gage, R.P. (1952) Survival Curve for Cancer Patients Following Treatment. *Journal of the American Statistical Association*, **47**, 501-515. <https://doi.org/10.1080/01621459.1952.10501187>
- [3] Klebanov, L.B., Rachev, S.T. and Yakovlev, A.Y. (1993) A Stochastic Model of Radiation Carcinogenesis: Latent Time Distributions and Their Properties. *Mathematical Biosciences*, **113**, 51-75. [https://doi.org/10.1016/0025-5564\(93\)90008-x](https://doi.org/10.1016/0025-5564(93)90008-x)
- [4] Chen, M., Ibrahim, J.G. and Sinha, D. (1999) A New Bayesian Model for Survival Data with a Surviving Fraction. *Journal of the American Statistical Association*, **94**, 909-919. <https://doi.org/10.1080/01621459.1999.10474196>
- [5] Chen, M., Ibrahim, J.G. and Sinha, D. (2002) Bayesian Inference for Multivariate Survival Data with a Cure Fraction. *Journal of Multivariate Analysis*, **80**, 101-126. <https://doi.org/10.1006/jmva.2000.1975>
- [6] Tsodikov, A. (2002) Semi-Parametric Models of Long- and Short-Term Survival: An Application to the Analysis of Breast Cancer Survival in Utah by Age and Stage. *Statistics in Medicine*, **21**, 895-920. <https://doi.org/10.1002/sim.1054>
- [7] Tsodikov, A.D., Ibrahim, J.G. and Yakovlev, A.Y. (2003) Estimating Cure Rates from Survival Data. *Journal of the American Statistical Association*, **98**, 1063-1078. <https://doi.org/10.1198/01622145030000001007>
- [8] Kutal, D.H. and Qian, L. (2018) A Non-Mixture Cure Model for Right-Censored Data with Fréchet Distribution. *Stats*, **1**, 176-188. <https://doi.org/10.3390/stats1010013>
- [9] Matthews, D.E. and Farewell, V.T. (1982) On Testing for a Constant Hazard against a Change-Point Alternative. *Biometrics*, **38**, 463-468. <https://doi.org/10.2307/2530460>
- [10] Müller, H. and Wang, J. (1990) Nonparametric Analysis of Changes in Hazard Rates for Censored Survival Data: An Alternative to Change-Point Models. *Biometrika*, **77**, 305-314. <https://doi.org/10.1093/biomet/77.2.305>
- [11] Pons, O. (2003) Estimation in a Cox Regression Model with a Change-Point According to a Threshold in a Covariate. *The Annals of Statistics*, **31**, 442-463. <https://doi.org/10.1214/aos/1051027876>
- [12] Dupuy, J. (2006) Estimation in a Change-Point Hazard Regression Model. *Statistics & Probability Letters*, **76**, 182-190. <https://doi.org/10.1016/j.spl.2005.07.013>
- [13] Li, Y., Qian, L. and Zhang, W. (2013) Estimation in a Change-Point Hazard Regression Model with Long-Term Survivors. *Statistics & Probability Letters*, **83**, 1683-1691. <https://doi.org/10.1016/j.spl.2013.03.026>
- [14] Zhao, X.B., et al. (2009) A Change-Point Model for Survival Data with Long-Term Survivors. *Stat Sinica*, **19**, 377-390.
- [15] Othus, M., Li, Y. and Tiwari, R. (2012) Change-Point Cure Models with Application to Estimating the Change-Point Effect of Age of Diagnosis among Prostate Cancer Patients. *Journal of Applied Statistics*, **39**, 901-911. <https://doi.org/10.1080/02664763.2011.626849>
- [16] Taweab, F., Ibrahim, N.A. and Arasan, J. (2015) A Bounded Cumulative Hazard Model with a Change-Point According to a Threshold in a Covariate for Right-Censored Data. *Applied Mathematics & Information Sciences*, **9**, 69-74. <https://doi.org/10.12785/amis/090110>
- [17] Kutal, D.H. (2018) Parameter Estimation in Mixture and Non-Mixture Cure Models. Ph.D. Thesis.

Supplement A: Derivatives of t_{ji} with respect to a_j , b_j , and τ

For subject i and regime $j \in \{1, 2\}$, define

$$t_{ji} = \delta_i \ln\{-\ln(p_j)\} + \delta_i \ln f_j(y_i) + F_j(y_i) \ln(p_j),$$

where the uncured-time distribution is exponential with

$$f_j(y_i) = \mu_j e^{-\mu_j y_i}, \quad F_j(y_i) = 1 - e^{-\mu_j y_i},$$

and $\delta_i \in \{0, 1\}$ is the censoring indicator. We use unconstrained parameters

$$a_j = \text{logit}(p_j) \Rightarrow p_j = \frac{1}{1 + e^{-a_j}}, \quad b_j = \ln(\mu_j) \Rightarrow \mu_j = e^{b_j}.$$

A.1: Derivatives with respect to a_j (via p_j)

Now t_{ji} depends on p_j only through $\ln\{-\ln(p_j)\}$ and $\ln(p_j)$ (weighted by $F_j(y_i)$). The derivatives with respect to p_j are

$$\begin{aligned} \frac{\partial}{\partial p_j} \ln\{-\ln(p_j)\} &= \frac{1}{p_j \ln(p_j)}, \quad \frac{\partial^2}{\partial p_j^2} \ln\{-\ln(p_j)\} = -\frac{\ln(p_j) + 1}{p_j^2 \ln^2(p_j)}, \\ \frac{\partial}{\partial p_j} \ln(p_j) &= \frac{1}{p_j}, \quad \frac{\partial^2}{\partial p_j^2} \ln(p_j) = -\frac{1}{p_j^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\partial t_{ji}}{\partial p_j} &= \delta_i \left(\frac{1}{p_j \ln(p_j)} \right) + \frac{F_j(y_i)}{p_j}, \\ \frac{\partial^2 t_{ji}}{\partial p_j^2} &= \delta_i \left(-\frac{\ln(p_j) + 1}{p_j^2 \ln^2(p_j)} \right) - \frac{F_j(y_i)}{p_j^2}. \end{aligned}$$

Chain rule to a_j gives

$$\frac{\partial p_j}{\partial a_j} = p_j(1 - p_j), \quad \frac{\partial^2 p_j}{\partial a_j^2} = p_j(1 - p_j)(1 - 2p_j).$$

Thus,

$$\begin{aligned} \frac{\partial t_{ji}}{\partial a_j} &= \frac{\partial t_{ji}}{\partial p_j} \frac{\partial p_j}{\partial a_j} = \left(\frac{\delta_i}{p_j \ln(p_j)} + \frac{F_j(y_i)}{p_j} \right) p_j(1 - p_j), \\ \frac{\partial^2 t_{ji}}{\partial a_j^2} &= \left(\frac{\partial^2 t_{ji}}{\partial p_j^2} \right) [p_j(1 - p_j)]^2 + \left(\frac{\partial t_{ji}}{\partial p_j} \right) p_j(1 - p_j)(1 - 2p_j). \end{aligned}$$

A.2: Derivatives with Respect to b_j (via μ_j)

For the exponential model,

$$\frac{\partial \ln f_j(y_i)}{\partial \mu_j} = \frac{1}{\mu_j} - y_i, \quad \frac{\partial \ln F_j(y_i)}{\partial \mu_j} = \frac{y_i e^{-\mu_j y_i}}{F_j(y_i)}.$$

Hence

$$\frac{\partial t_{ji}}{\partial \mu_j} = \delta_i \left(\frac{1}{\mu_j} - y_i \right) + \ln(p_j) \frac{\partial F_j(y_i)}{\partial \mu_j}, \quad \frac{\partial F_j(y_i)}{\partial \mu_j} = y_i e^{-\mu_j y_i}.$$

So

$$\frac{\partial t_{ji}}{\partial \mu_j} = \delta_i \left(\frac{1}{\mu_j} - y_i \right) + \ln(p_j) y_i e^{-\mu_j y_i}.$$

For the second derivative,

$$\frac{\partial^2 t_{ji}}{\partial \mu_j^2} = -\frac{\delta_i}{\mu_j^2} - \ln(p_j) y_i^2 e^{-\mu_j y_i}.$$

Chain to $b_j = \ln \mu_j$ with $\partial \mu_j / \partial b_j = \mu_j$, $\partial^2 \mu_j / \partial b_j^2 = \mu_j$, we obtain

$$\frac{\partial t_{ji}}{\partial b_j} = \mu_j \left[\delta_i \left(\frac{1}{\mu_j} - y_i \right) + \ln(p_j) y_i e^{-\mu_j y_i} \right] = \delta_i (1 - \mu_j y_i) + \mu_j \ln(p_j) y_i e^{-\mu_j y_i},$$

$$\frac{\partial^2 t_{ji}}{\partial b_j^2} = \mu_j^2 \left(-\frac{\delta_i}{\mu_j^2} - \ln(p_j) y_i^2 e^{-\mu_j y_i} \right) + \mu_j \left[\delta_i \left(\frac{1}{\mu_j} - y_i \right) + \ln(p_j) y_i e^{-\mu_j y_i} \right].$$

A.3: Derivatives with Respect to τ

The derivative of the smoothed log-likelihood with respect to the change point τ is

$$\frac{\partial \ell^*}{\partial \tau} = \sum_{i=1}^n \frac{\partial k_n(\tau - x_i)}{\partial \tau} (t_{1i} - t_{2i}) = \frac{1}{h_n} \sum_{i=1}^n k_n(\tau - x_i) [1 - k_n(\tau - x_i)] (t_{1i} - t_{2i}).$$

These expressions provide the analytic score contributions and observed-information terms needed for Newton–Raphson estimation of $(p_1, p_2, \mu_1, \mu_2, \tau)$ in the smoothed-likelihood framework.

A.4: Compact Score (Gradient) for the Smoothed Log-Likelihood

Smoothing weights.

For each subject i , define the logistic weight

$$k_i = \frac{\exp\{(\tau - x_i)/h_n\}}{1 + \exp\{(\tau - x_i)/h_n\}} \in (0, 1), \quad 1 - k_i = \frac{1}{1 + \exp\{(\tau - x_i)/h_n\}}.$$

Let $F_j(y_i) = 1 - e^{-\mu_j y_i}$ and $e_j(y_i) = e^{-\mu_j y_i}$ for $j \in \{1, 2\}$, and recall

$$a_j = \text{logit}(p_j) \Rightarrow p_j = \frac{1}{1 + e^{-a_j}}, \quad b_j = \ln(\mu_j) \Rightarrow \mu_j = e^{b_j}.$$

The per-subject regime contributions are

$$t_{ji} = \delta_i \ln\{-\ln(p_j)\} + \delta_i (\ln \mu_j - \mu_j y_i) + \ln(p_j) F_j(y_i), \quad j = 1, 2.$$

Score (gradient).

The smoothed log-likelihood is

$$\ell^*(\theta) = \sum_{i=1}^n \{k_i t_{1i} + (1-k_i) t_{2i}\}.$$

Its gradient $g(\eta) = \nabla_{\eta} \ell^*(\eta)$ with $\eta = (a_1, a_2, b_1, b_2, \tau)^\top$ is

$$\frac{\partial \ell^*}{\partial a_1} = \sum_{i=1}^n k_i (1-p_1) \left(\frac{\delta_i}{p_1 \ln p_1} + \frac{F_1(y_i)}{p_1} \right)$$

$$\frac{\partial \ell^*}{\partial a_2} = \sum_{i=1}^n (1-k_i) (1-p_2) \left(\frac{\delta_i}{p_2 \ln p_2} + \frac{F_2(y_i)}{p_2} \right)$$

$$\frac{\partial \ell^*}{\partial b_1} = \sum_{i=1}^n k_i \left[\delta_i (1 - \mu_1 y_i) + \mu_1 (\ln p_1) y_i e^{-\mu_1 y_i} \right]$$

$$\frac{\partial \ell^*}{\partial b_2} = \sum_{i=1}^n (1-k_i) \left[\delta_i (1 - \mu_2 y_i) + \mu_2 (\ln p_2) y_i e^{-\mu_2 y_i} \right]$$

$$\frac{\partial \ell^*}{\partial \tau} = \frac{1}{h_n} \sum_{i=1}^n k_i (1-k_i) (t_{1i} - t_{2i}).$$

A.5: Observed Information/Hessian

Building-block derivatives (per subject i , regime j).

For compactness, let

$$A_{ji} = \frac{\partial^2 t_{ji}}{\partial a_j^2}, \quad B_{ji} = \frac{\partial^2 t_{ji}}{\partial b_j^2}, \quad C_{ji} = \frac{\partial^2 t_{ji}}{\partial a_j \partial b_j}.$$

These follow directly from the scalar derivations via chain rule.

(i) *Second derivative in a_j .* With $p_j = \text{logit}^{-1}(a_j)$ and

$$\frac{\partial t_{ji}}{\partial p_j} = \frac{\delta_i}{p_j \ln p_j} + \frac{F_j(y_i)}{p_j}, \quad \frac{\partial^2 t_{ji}}{\partial p_j^2} = -\frac{\delta_i (\ln p_j + 1)}{p_j^2 \ln^2 p_j} - \frac{F_j(y_i)}{p_j^2},$$

and

$$\frac{\partial p_j}{\partial a_j} = p_j (1-p_j), \quad \frac{\partial^2 p_j}{\partial a_j^2} = p_j (1-p_j) (1-2p_j),$$

we have

$$A_{ji} = \left(\frac{\partial^2 t_{ji}}{\partial p_j^2} \right) \left[p_j (1-p_j) \right]^2 + \left(\frac{\partial t_{ji}}{\partial p_j} \right) p_j (1-p_j) (1-2p_j).$$

(ii) *Second derivative in b_j .* Using $\mu_j = e^{b_j}$,

$$\frac{\partial t_{ji}}{\partial b_j} = \delta_i (1 - \mu_j y_i) + \mu_j (\ln p_j) y_i e^{-\mu_j y_i},$$

$$B_{ji} = \frac{\partial^2 t_{ji}}{\partial b_j^2} = \mu_j^2 \left(-\frac{\delta_i}{\mu_j^2} - (\ln p_j) y_i^2 e^{-\mu_j y_i} \right) + \mu_j \left(\frac{\delta_i}{\mu_j} - \delta_i y_i + (\ln p_j) y_i e^{-\mu_j y_i} \right).$$

(You may leave B_{ji} in this equivalent unsimplified form; it expands directly from the chain rule with $\partial \mu_j / \partial b_j = \mu_j$, $\partial^2 \mu_j / \partial b_j^2 = \mu_j$.)

(iii) *Mixed derivative in (a_j, b_j)* . Only the term $\ln(p_j)F_j(y_i)$ couples p_j and μ_j . From either order,

$$C_{ji} = \frac{\partial^2 t_{ji}}{\partial a_j \partial b_j} = (1 - p_j) \mu_j y_i e^{-\mu_j y_i}.$$

Hessian entries (sum over subjects).

All cross-regime blocks are zero because t_{1i} depends only on (a_1, b_1) and t_{2i} only on (a_2, b_2) .

$$\frac{\partial^2 \ell^*}{\partial a_1^2} = \sum_{i=1}^n k_i A_{1i}, \quad \frac{\partial^2 \ell^*}{\partial a_2^2} = \sum_{i=1}^n (1 - k_i) A_{2i}$$

$$\frac{\partial^2 \ell^*}{\partial b_1^2} = \sum_{i=1}^n k_i B_{1i}, \quad \frac{\partial^2 \ell^*}{\partial b_2^2} = \sum_{i=1}^n (1 - k_i) B_{2i}$$

$$\frac{\partial^2 \ell^*}{\partial a_1 \partial b_1} = \sum_{i=1}^n k_i C_{1i}, \quad \frac{\partial^2 \ell^*}{\partial a_2 \partial b_2} = \sum_{i=1}^n (1 - k_i) C_{2i}$$

$$\frac{\partial^2 \ell^*}{\partial a_1 \partial a_2} = \frac{\partial^2 \ell^*}{\partial a_1 \partial b_2} = \frac{\partial^2 \ell^*}{\partial b_1 \partial a_2} = \frac{\partial^2 \ell^*}{\partial b_1 \partial b_2} = 0$$

Hessian entries involving τ .

Since t_{1i} and t_{2i} do not depend on τ ,

$$\frac{\partial^2 \ell^*}{\partial \tau^2} = \frac{1}{h_n^2} \sum_{i=1}^n k_i (1 - k_i) (1 - 2k_i) (t_{1i} - t_{2i})$$

and for $\theta \in \{a_1, b_1\}$,

$$\begin{aligned} \frac{\partial^2 \ell^*}{\partial \tau \partial \theta} &= \frac{1}{h_n} \sum_{i=1}^n k_i (1 - k_i) \frac{\partial t_{1i}}{\partial \theta}, \\ \text{where } \frac{\partial t_{1i}}{\partial a_1} &= (1 - p_1) \left(\frac{\delta_i}{p_1 \ln p_1} + \frac{F_1(y_i)}{p_1} \right), \\ \text{and } \frac{\partial t_{1i}}{\partial b_1} &= \delta_i (1 - \mu_1 y_i) + \mu_1 (\ln p_1) y_i e^{-\mu_1 y_i}. \end{aligned}$$

and for $\theta \in \{a_2, b_2\}$,

$$\begin{aligned} \frac{\partial^2 \ell^*}{\partial \tau \partial \theta} &= -\frac{1}{h_n} \sum_{i=1}^n k_i (1 - k_i) \frac{\partial t_{2i}}{\partial \theta}, \\ \text{where } \frac{\partial t_{2i}}{\partial a_2} &= (1 - p_2) \left(\frac{\delta_i}{p_2 \ln p_2} + \frac{F_2(y_i)}{p_2} \right), \\ \text{and } \frac{\partial t_{2i}}{\partial b_2} &= \delta_i (1 - \mu_2 y_i) + \mu_2 (\ln p_2) y_i e^{-\mu_2 y_i}. \end{aligned}$$

Implementation notes.

1) The blocks are *observed* second derivatives; the negative of the Hessian is the observed information. 2) The mixed blocks across regimes vanish, simplifying Newton-Raphson updates. 3) The formulas are numerically stable provided $p_j \in (0, 1)$ and $\ln p_j < 0$; guard divisions by $\ln p_j$.