The New England Journal of Statistics in Data Science Volume 3, 28–41 (2025)
DOI: https://doi.org/10.51387/24-NEJSDS70

Bayesian Inference of A Unified Estimand under Survival Models with Cure Fraction

Hongfei LI, Qian H. LI, and Ming-Hui CHEN*

Abstract

Cure models are gaining more and more popularity for modeling time-to-event data for different forms of cancer, for which a considerable proportion of patients are considered "cured." Two types of cure models are widely used, the mixture cure model (MCM) and the promotion time cure model (PTCM). In this article, we propose a unified estimand Δ for comparing treatment and control groups under the survival models with cure fraction, which focuses on whether the treatment extends survival for patients. In addition, we introduce a general framework of Bayesian inference under the cure models. Simulation studies demonstrate that regardless of whether the model is correctly specified, the inference of the unified estimand Δ yields desirable empirical performance. We analyze the ECOG's melanoma cancer data E1684 via the unified estimand Δ under different models to further demonstrate the proposed methodology.

keywords and phrases: Cure model, E1684, MCMC, Bayesian hypothesis testing, Covariate adjustment, G-computation.

1. INTRODUCTION

Modeling time-to-event data for cancer treatments has become an increasingly important area of research. In recent years, survival models with cure fraction, also called cure models, have gained increasing interest, in which a considerable proportion of patients are considered "cured," that is, to remain disease-free after a certain follow-up period. It has been used for different forms of cancer such as breast cancer, non-Hodgkins lymphoma, leukemia, prostate cancer, melanoma, and head and neck cancer [13].

Two types of cure models, the mixture cure model (MCM) and the promotion time cure model (PTCM), are commonly used. The mixture cure model (MCM), proposed by Berkson and Gage [1], is formed by two mixture components, and it has been commonly used and discussed in [7, 8, 12, 23]. Other than the standard mixture model for cure rates, Chen et al. [3] proposed a promotion time cure model (PTCM) based on tumor growth characteristics using the Poisson distribution. These models are built on different assumptions, challenging researchers to choose the appropriate model for their study and interpret the different definitions of the treatment effect. The connection between the two types of models was discussed in [3], and Yin and Ibrahim [24] proposed a unified approach to bridge the mixture cure model and the promotion cure model via Box-Cox transformation. However, with different choices of the power parameter, the interpretations of the parameters are different and, thus, it is not straightforward to define the treatment effect. Therefore, a unified approach is needed to decide between these two types of cure models and to evaluate the treatment effect with different model choices.

In 2021, the Food and Drug Administration (FDA) released a guidance entitled "E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials,"[10] which introduced the concept of "Estimand." The addendum provides a structured estimand framework to enhance the communication amongst disciplines about the clinical trial objectives, design, conduct, analysis, and interpretation. Also, it emphasizes the attention regarding the treatment effects of interest that a clinical trial should investigate, which properly informs decision-making for the pharmaceutical industry. Furthermore, FDA published a final guidance in 2023 entitled "Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products,"[11] which mainly focuses on the use of prognostic covariates to improve statistical efficiency for estimating and testing treatment effects. It points out that adjusting for prognostic covariates in the analyses of efficacy endpoints in randomized clinical trials will generally reduce the variability of estimation of treatment effects and thus lead to narrower confidence intervals and power gains in hypothesis testing. Specifically, it emphasizes the need to differentiate conditional and unconditional treatment effects to ensure the specific objectives defined within the "Estimand" framework are examined when making statistical inferences. It should be noted that the conditional treatment effects, such as the hazard ratio in survival models, are influenced by choices in model and variable selections, posing challenges to the intentions of the invariant "Estimand" concept. Therefore, the guidance and the topics of "Estimand"

*Corresponding author.

and "Covariate Adjustment" are particularly relevant to the discussion of treatment effects in survival models with cure fraction.

This article sets out with two primary objectives. Firstly, we propose a unified estimand for assessing treatment effects under survival models with cure fraction that is invariant to model and variable selections and examine its theoretical properties. Secondly, we develop a Bayesian inference framework for the unified estimand, and illustrate its usefulness through simulation and a case study.

The rest of the paper is organized as follows. Section 2 provides a summary of the survival models with cure fraction. Section 3 introduces the proposed unified estimand under survival models with cure fraction. The Bayesian inference of the treatment effects is discussed in Section 4, including Bayesian hypothesis testing, model comparisons, and Bayesian computation. Section 5 evaluates the performance of the unified estimand through a comprehensive simulation study. An application of the proposed methodology to the E1684 melanoma cancer data is presented in Section 6. Section 7 concludes the paper with a discussion of main findings with future research directions.

2. SURVIVAL MODELS WITH CURE FRACTION

2.1 Mixture Cure Models

The concept of the mixture cure models (MCM) was first introduced by Berkson and Gage [1], in which it is assumed that a certain fraction π of the population is cured ($y = 1$) and the remaining $1 - \pi$ is not cured ($y = 0$). For the i-th individual, y_i denotes the unobserved cured indicator, and $\pi_i = \mathbb{P}(y_i = 1) \in (0, 1)$ denotes the cured probability. If a patient is cured, it is assumed that the patient will not die for a sufficiently long period of time. The survival (failure) time t_i^{surv} can be written as

$$t_i^{surv} = \begin{cases} t_i^{nc} & \text{if } y_i = 0 \\ \infty & \text{if } y_i = 1 \end{cases},$$

where t_i^{nc} is the survival time of subject i if the individual is not cured. Let $t_i = \min\{t_i^{surv}, c_i\}$ denote the observed right-censored survival time, where c_i denotes the censoring time and $\delta_i = 1\{t_i^{surv} < c_i\}$ denotes the censoring indicator for subject i such that $\delta_i = 1$ if t_i is a failure time and 0 if t_i is right-censored. If subject i is not cured, i.e., $y_i = 0$, denote the probability density function, the survival function, and hazard function for the failure time t_i^{nc} by $f_i^{nc}(t)$, $S_i^{nc}(t)$ and $h_i^{nc}(t)$.

Since the cure indicator y_i is unobserved, considering the existence of cure fraction, the unconditional probability density function of the survival time for the i-th individual can be expressed as

$$f_i(t) = (1 - \pi_i) f_i^{nc}(t),$$

and the unconditional survival function is given by

$$S_i(t) = \mathbb{P}(T_i > t) = \pi_i + (1 - \pi_i) S_i^{nc}(t). \tag{2.1}$$

The hazard function for the mixture cure model can be derived as

$$h_i(t) = \frac{(1 - \pi_i) f_i^{nc}(t)}{\pi_i + (1 - \pi_i) S_i^{nc}(t)}. \tag{2.2}$$

Since $\lim_{t\to\infty} S_i^{nc}(t) = 0$, $\lim_{t\to\infty} S_i(t) = \pi_i > 0$, the survival function for the mixture cure model is not a proper survival function. The hazard ratio between subject i and i' is a function of time t. Thus, the mixture cure model is a non-proportional hazards model. When $\pi_i \to 0$, $S_i(t) \to S_i^{nc}(t)$ and $h_i(t) \to h_i^{nc}(t)$, the mixture cure model reduces to the standard survival model.

2.2 Promotion Time Cure Model

The promotion time cure model (PTCM) was developed by Chen et al. [3], where a Poisson distribution is assumed for the number of active carcinogenic cells. For the i-th subject, let N_i denote the number of active carcinogenic cells, and assume that $N_i \sim \mathcal{P}(\lambda_i)$ follows a Poisson distribution with mean λ_i. By assuming the distribution of N_i, the cured probability can be expressed as

$$\pi_i = \mathbb{P}(N_i = 0) = \exp(-\lambda_i).$$

Denote $Z_{ij}, j = 1, \ldots, N_i$ as i.i.d random variables of incubation time for the j-th carcinogenic cells. It is assumed that Z_{ij} follows the same distribution independent of N_i, and let $f(t)$ and $S(t)$ denote its density function and survival function. The event time for i-th individual is defined as $T_i = \min\{Z_{ij}, 0 \le j \le N_i\}$, where $P(Z_{i0} = \infty) = 1$.

The survival function for the i-th individual can be expressed as

$$\begin{aligned} S_i(t) &= \sum_{k=0}^{\infty} \mathbb{P}(N_i = k)\mathbb{P}(Z_{ij} > t, j = 0, \ldots, k) \\ &= \sum_{k=0}^{\infty} \frac{\lambda_i^k}{k!} \exp(-\lambda_i) S(t)^k \\ &= \exp(-\lambda_i(1 - S(t))). \end{aligned} \tag{2.3}$$

Since $\lim_{t\to\infty} S(t) = 0$, $\lim_{t\to\infty} S_i(t) = \exp(-\lambda_i) > 0$. Therefore, similar to the MCM, the survival function for the subject in the PTCM $S_i(t)$ is also not a proper survival function. The density function and hazard function of the survival time for the i-th subject can be further expressed as

$$f_i(t) = \lambda_i f(t) \exp(-\lambda_i(1 - S(t))), \tag{2.4}$$

and

$$h_i(t) = \lambda_i f(t) \tag{2.5}$$

Under this promotion time cure model, the hazard ratio between i-th subject and i'-th subject is denoted as

$$\frac{h_i(t)}{h_{i'}(t)} = \frac{\lambda_i}{\lambda_{i'}},$$

which is a constant over time. Thus, the promotion time cure model follows the proportional hazards assumption.

2.3 Regression Forms and Likelihood Functions

2.3.1 Mixture Cure Model

We consider the mixture cure model defined in Section 2.1. Since both the cure fraction component and the non-cured survival component can be affected by a set of prognostic covariates, these two components can be modeled separately in regression forms. For the i-th subject, let $\boldsymbol{x}_i = (1, z_i, \tilde{\boldsymbol{x}}_i^\top)^\top$ denote a $(p+2)$-dimensional vector of covariates, where $\tilde{\boldsymbol{x}}_i$ is the prognostic factors and z_i is the treatment group indicator such that $z_i = 1$ if the i-th subject is in the treatment group and $z_i = 0$ if the subject is assigned to the control group. We assume a logistic regression model for y_i given by

$$\pi_i = \mathbb{P}(y_i = 1) = \sigma(-\boldsymbol{x}_i^\top \boldsymbol{\beta}) = \frac{1}{1 + \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}, \tag{2.6}$$

where $\sigma(.)$ denotes the standard logistic function, $\boldsymbol{\beta} = (\beta_0, \beta_z, \tilde{\boldsymbol{\beta}}^\top)^\top \in \mathbb{R}^{p+2}$ is a $(p+2)$-dimensional vector of regression coefficients. Especially, β_0, β_z, and $\tilde{\boldsymbol{\beta}}$ are the regression coefficients associated with the intercept, the treatment indicators z_i, and $\tilde{\boldsymbol{x}}_i$. After adjusting for the other covariates, β_z represents the conditional treatment effect in the log odds ratio of being cured between the treatment and control groups. For the non-cured survival component, to cover a wide variety of survival curves while maintaining the proportional hazards assumption, the Weibull distributed failure time is assumed. If the i-th individual is not cured, the survival function is assumed to be

$$S_i^{nc}(t \mid \boldsymbol{x}_i) = \exp(-t^\alpha \exp(\boldsymbol{x}_i^\top \boldsymbol{\gamma})), \tag{2.7}$$

where α is a fixed but unknown Weibull shape parameter, $\boldsymbol{\gamma} = (\gamma_0, \gamma_z, \tilde{\boldsymbol{\gamma}}^\top)^\top \in \mathbb{R}^{p+2}$ is the vector of regression coefficients associated with $\boldsymbol{x}_i$ for the non-cured survival component, and $\exp(\boldsymbol{x}_i^\top \boldsymbol{\gamma})$ is the subject-specific scale parameter. After adjusting for the other covariates, γ_z represents the conditional treatment effect in the log hazard ratio between the treatment and control groups for the non-cured patients.

Plugging (2.6) and (2.7) in (2.1), the unconditional survival function for the mixture cure model is given by

$$\begin{aligned} S_i(t \mid \boldsymbol{x}_i) = & \frac{1}{1 + \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})} + \\ & \frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})} \exp(-t^\alpha \exp(\boldsymbol{x}_i^\top \boldsymbol{\gamma})). \end{aligned} \tag{2.8}$$

Let $\boldsymbol{\theta}_M = \{\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha\}$ denote the set of parameters, where $\{\beta_z, \gamma_z\}$ serve as the conditional treatment effect parameters of interest. Suppose there are n subjects in the dataset. Let $\mathcal{D}^{obs} = \{\boldsymbol{x}_i, t_i, \delta_i, i = 1, \ldots, n\}$ denote the observed data under the mixture cure model, $\mathcal{Y} = \{y_i, i = 1, \ldots, n\}$ denote the hidden state of cured indicators, and let $\mathcal{D}_M^{comp} = \mathcal{D}^{obs} \cup \mathcal{Y}$ denote the complete data.

For the individual with failure time ($\delta_i = 1$), the likelihood can be expressed by multiplying the non-cured probability $1 - \pi_i$ with the probability density function of non-cured patient $f_i^{nc}(t)$. For the censored individual with $\delta_i = 0$, the likelihood can be constructed by adding the cured probability with the probability of non-cured survival. The observed-data likelihood function is given by

$$\begin{aligned} \mathcal{L}_M^{obs}(\boldsymbol{\theta}_M \mid \mathcal{D}^{obs}) = & \prod_{\delta_i = 1} \Big((1 - \pi_i) f_i^{nc}(t_i \mid \boldsymbol{x}_i) \Big) \times \\ & \prod_{\delta_i = 0} \Big(\pi_i + (1 - \pi_i) S_i^{nc}(t_i \mid \boldsymbol{x}_i) \Big). \end{aligned} \tag{2.9}$$

In the second parenthesis of the observed-data likelihood, the mixture component can be converted into a multiplied form after including the hidden state $\mathcal{Y}$ in the likelihood function. The complete-data likelihood can be written as

$$\begin{aligned} \mathcal{L}_M^{comp}(\boldsymbol{\theta}_M \mid \mathcal{D}_M^{comp}) = & \prod_{y_i = 1} \pi_i \times \\ & \prod_{y_i = 0} \Big((1 - \pi_i) f_i^{nc}(t_i \mid \boldsymbol{x}_i)^{\delta_i} S_i^{nc}(t_i \mid \boldsymbol{x}_i)^{(1 - \delta_i)} \Big). \end{aligned} \tag{2.10}$$

Note that under mixture cure model, the cured fraction component (2.6) and non-cured survival component (2.7) are linked through the hidden state $\mathcal{Y}$. The following Theorem and Remarks imply that conditioning on $\mathcal{D}^{obs}$, the key parameters of interest β_z and γ_z interact with each other. Also, the maximum-likelihood estimators (MLE) and the posterior samples of β_z and γ_z are not independent.

Theorem 1. *Under the mixture cure regression model defined in (2.8), the (β_z, γ_z) entry of the observed Fisher information*

$$\mathcal{I}(\boldsymbol{\theta}_M)_{(\beta_z, \gamma_z)} = -\frac{\partial^2}{\partial \beta_z \partial \gamma_z} \ell_M^{obs}(\boldsymbol{\theta}_M \mid \mathcal{D}^{obs}) > 0.$$

The proof is given in Appendix A.

Remark 1. The off-diagonal entry in observed Fisher information matrix $\mathcal{I}(\boldsymbol{\theta}_M)_{(\beta_z, \gamma_z)} > 0$ likely implies the off-diagonal entry in the inverse Fisher information matrix $\mathcal{I}^{-1}(\boldsymbol{\theta}_M)_{(\beta_z, \gamma_z)} < 0$.

Intuitively, since the hidden state $\mathcal{Y}$ is unknown, the conditional log odds ratio estimates from the cure rate component and the conditional log hazard ratio estimates from the non-cured population are correlated. The following Remarks

reveal how the correlation will impact the maximum likelihood estimator (MLE) from the Frequentist perspective and posterior samples from Bayesian perspective.

Remark 2. Since the covariance matrix associated with the maximum-likelihood estimators can be approximated by the inverse matrix of the Fisher information matrix, it is likely that under mixture cure model,

$$\mathrm{Corr}(\widehat{\beta_z}, \widehat{\gamma_z}) < 0,$$

where $\widehat{\beta_z}$ and $\widehat{\gamma_z}$ refer to the MLEs of β_z and γ_z. That is, $\widehat{\beta_z}$ and $\widehat{\gamma_z}$ are negatively correlated.

Remark 3. By Bernstein–von Mises Theorem in Bayesian statistics, the asymptotic distribution of the posterior converges to a multivariate Gaussian distribution with covariance matrix given by $n^{-1}\mathcal{I}(\boldsymbol{\theta}_M)^{-1}$. Thus, the posterior samples of β_z and γ_z are correlated.

Considering the correlation, looking at the conditional treatment effects from the cured fraction component and the non-cured survival component separately for the mixture cure model is not ideal. A unified estimand (measure of treatment effect) should be proposed to unify the treatment effect measure and enable comparisons.

In one extreme case, when $\beta_0 \to \infty$, the cured probability π_i goes to 0 for all subjects, and the mixture cure model reduces to the Weibull regression model with survival function

$$S_i(t \mid \boldsymbol{x}_i) = \exp(-t^\alpha \exp(\boldsymbol{x}_i^\top \boldsymbol{\gamma})), \tag{2.11}$$

where α is a fixed but unknown Weibull shape parameter, $\boldsymbol{\gamma} = (\gamma_0, \gamma_z, \tilde{\boldsymbol{\gamma}}^\top)^\top \in \mathbb{R}^{p+2}$ is the vector of regression coefficients associated with $\boldsymbol{x}_i$ under Weibull regression model, and $\exp(\boldsymbol{x}_i^\top \boldsymbol{\gamma})$ is the subject-specific scale parameter. The model becomes proportional hazards, and γ_z represents the treatment effect in the log hazard ratio between the treatment and control groups after adjusting for other factors. The reduced model can be fitted by Cox regression by maximizing the partial likelihood function

$$\mathcal{PL}(\boldsymbol{\gamma}) = \prod_{j=1}^{d} \prod_{i \in D(t_{(j)})} \frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\gamma})}{\sum_{i \in R(t_{(j)})} \exp(\boldsymbol{x}_i^\top \boldsymbol{\gamma})}, \tag{2.12}$$

where $t_{(1)} < t_{(2)} < \cdots < t_{(d)}$ are the d ordered distinct event times, $D(t)$ denotes the set of subjects who die at time t, and $R(t)$ represents the risk set at time t^-, that is, the set of individuals who have not failed or been censored by that time. The inference with respect to $\boldsymbol{\gamma}$ can be made via the semi-parametric partial likelihood $\mathcal{PL}(\boldsymbol{\gamma})$.

2.3.2 Promotion Time Cure Model

For the promotion time cure model defined in (2.3), for the i-th subject, let

$$\lambda_i = \exp(\boldsymbol{x}_i^\top \boldsymbol{\zeta}), \tag{2.13}$$

where $\boldsymbol{\zeta} = (\zeta_0, \zeta_z, \tilde{\boldsymbol{\zeta}}^\top)^\top \in \mathbb{R}^{p+2}$ is a $(p+2)$-dimensional vector of regression coefficients. Especially, ζ_z is the regression coefficient associated with the treatment indicators z_i. After adjusting for the other covariates, ζ_z represents the conditional treatment effect in the log hazard ratio between the treatment and control groups under this model. The relationship serves as a canonical link under the Poisson regression model. Parameter ζ_z is the conditional treatment effect parameter of interest in the PTCM.

Assume the incubation time for each active carcinogenic cell follows an *i.i.d.* Weibull distribution with shape parameter α and scale parameter $\exp(\mu)$. The survival function for the incubation is given by

$$S(t) = \exp(-t^\alpha \exp(\mu)). \tag{2.14}$$

The survival function and probability density function for the promotion time cure model can be expressed as

$$S_i(t \mid \boldsymbol{x}_i) = \exp(-\exp(\boldsymbol{x}_i^\top \boldsymbol{\zeta})(1 - \exp(-t^\alpha \exp(\mu)))), \tag{2.15}$$

and

$$f_i(t \mid \boldsymbol{x}_i) = \alpha t^{\alpha-1} \exp(\boldsymbol{x}_i^\top \boldsymbol{\zeta} + \mu - t^\alpha \exp(\mu)) S_i(t). \tag{2.16}$$

Suppose there are n subjects in the dataset. Under the promotion time cure model defined in (2.15), let $\mathcal{D}^{obs} = \{\boldsymbol{x}_i, t_i, \delta_i, i = 1, \ldots, n\}$ denote the observed data under mixture cure model, $\mathcal{N} = \{N_i, i = 1, \ldots, n\}$ denote unobserved data for the number of active carcinogenic cells, and let $\mathcal{D}_P^{comp} = \mathcal{D}^{obs} \cup \mathcal{N}$ denote the complete data. Let $\boldsymbol{\theta}_P = \{\boldsymbol{\zeta}, \alpha, \mu\}$ denote the parameters associated with the promotion time cure model. The observed-data likelihood function is given by

$$\begin{aligned} &\mathcal{L}_P^{obs}(\boldsymbol{\theta}_P \mid \mathcal{D}^{obs}) \\ =& \prod_{i=1}^{n} f_i(t_i \mid \boldsymbol{x}_i)^{\delta_i} S_i(t_i \mid \boldsymbol{x}_i)^{(1-\delta_i)} \\ =& \prod_{i=1}^{n} \exp(-\exp(\boldsymbol{x}_i^\top \boldsymbol{\zeta})(1 - \exp(-t_i^\alpha \exp(\mu)))) \times \\ & \prod_{\delta_i=1} \alpha t_i^{\alpha-1} \exp(\boldsymbol{x}_i^\top \boldsymbol{\zeta} + \mu - t_i^\alpha \exp(\mu)). \end{aligned} \tag{2.17}$$

3. A UNIFIED ESTIMAND UNDER CURE MODELS

Under the proportional-hazards models, the conditional treatment effect in the hazard ratio between treatments is overwhelmingly used to characterize efficacy of the treatment. However, as shown in (2.2), the mixture cure model does not belong to the proportional hazards framework. Thus, how to develop an appropriate measure of the treatment effect (estimand) for mixture cure models has become an interesting topic. There are various approaches to summarise and make inference of the treatment effects under

the non-proportional hazards models. The Weighted Log-Rank Test (WLRT) is one of the most popular statistical tests dealing with the non-proportional hazards models, and Fleming and Harrington [9] developed a class of test statistics $FH(\rho, \gamma)$. Another popular type of test is developed based on the Kaplan-Meier curve, including the Weighted Kaplan-Meier tests (WKM) [20], the restricted mean survival time (RMST) comparison [21], and a divergence measure between the two survival functions [6]. Especially, the difference in restricted mean survival time (Δ_{RMST}) shows an increasing popularity as an alternative to hazard ratio, and it is defined as

$$\begin{aligned}\Delta_{RMST}(t^*) =& \mathbb{E}[\min(T_1, t^*) - \min(T_0, t^*)] \\ =& \int_0^{t^*} [S_{pop,1}(t) - S_{pop,0}(t)]dt,\end{aligned} \tag{3.1}$$

where T_0 and T_1 denote the survival times for the control arm and the treatment arm, $S_{pop,1}(t)$ and $S_{pop,0}(t)$ represent the survival functions for the treatment and control arms, and t^* is the duration of follow-up or truncation time. $\Delta_{RMST}(t^*)$ describes differences between treatment and control arms in their t^*-year life expectancy. Moreover, several classes of combination tests have been developed in recent years, including the Breslow test [2], Lee's combo test [16], and the MaxCombo test [17]. Lin et al. [19] evaluated different tests for the non-proportional hazards models, and Li et al. [18] developed an overlapping approach to measure the comparability between two arms through resampling technique. Regarding the treatment effect measure, Chen et al. [5] developed the averaged hazard ratio estimates for the non-proportional hazards model.

Motivated by the restricted mean survival time (RMST) [21] and a divergence measure proposed in [6], we define a new unified estimand under the survival models of the form

$$\begin{aligned}\Delta &= \mathbb{E}[sgn(T_1 - T_0)] \\ &= \mathbb{P}(T_1 > T_0) - \mathbb{P}(T_0 > T_1) \\ &= \int_0^{\infty} (S_{pop,1}(t) f_{pop,0}(t) - S_{pop,0}(t) f_{pop,1}(t))dt \\ &\in [-1, 1],\end{aligned} \tag{3.2}$$

where T_0 and T_1 denote the survival times for the control arm and the treatment arm, $S_{pop,1}(t)$, $S_{pop,0}(t)$, $f_{pop,0}(t)$, and $f_{pop,1}(t)$ represent the survival functions and probability density functions for the treatment and control arms, and $sgn(.)$ is the sign function such that

$$sgn(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

Following the idea of the causal effect model, T_0 and T_1 can be viewed as the potential outcomes (survival time) if the individuals are assigned to the treatment arm and the control arm, respectively. An intuitive interpretation for Δ is that Δ characterizes the probability that the treatment extends survival for patients.

The proposed unified estimand, representative of the unconditional treatment effect between treatment and control groups, remains invariant to model and variable selections. This characteristic ensures its alignment with the FDA's guidelines and discussions about the "Estimand" concept.

The following Theorems give parametric derivations for the unified estimand Δ under the proportional hazards model, mixture cure model, and promotion time cure model. In addition, it connects the unconditional treatment effect in the proposed estimand with the conditional treatment effect under different models.

Theorem 2. *Under the proportional hazards model, i.e., $h_{pop,1}(t) = \exp(\gamma_z) h_{pop,0}(t)$, the unified estimand is a one-to-one transformation of the log hazard ratio γ_z.*

$$\Delta_{PH} = \frac{1 - \exp(\gamma_z)}{1 + \exp(\gamma_z)} = \tanh(-\frac{\gamma_z}{2}).$$

See proof in Appendix A.

Theorem 2 shows that the unified estimand Δ_{PH} serves as a one-to-one transformation, negative hyperbolic tangent transformation, of half of the log hazard ratio γ_z. When the treatment benefits patients ($\gamma_z < 0$), $\Delta_{PH} > 0$, and vice versa. Δ_{PH} quantifies the ratio between the differences in hazard functions to the sum of hazard functions.

Remark 4. Consider the Maclaurin Series expansion of $\tanh(x)$, we have

$$\frac{\Delta_{PH}}{\gamma_z} = \frac{\tanh(-\frac{\gamma_z}{2})}{\gamma_z} = -\frac{1}{2} + o(\gamma_z). \tag{3.3}$$

Theorem 3. *Under the mixture cure model defined in (2.8), when the prognostic variables are available, the unified estimand Δ_M is defined as the average treatment effect (ATE) based on the finite-sample population by integrating out the covariates, such that*

$$\begin{aligned}\Delta_M &= \mathbb{E}_{\tilde{\boldsymbol{X}}}[\mathbb{E}[sgn(T_1 - T_0) \mid \tilde{\boldsymbol{X}}]] \\ &= \pi_{01} \tanh(-\frac{\gamma_z}{2}) + \pi_1 - \pi_0,\end{aligned} \tag{3.4}$$

where $\pi_{01} = \mathbb{P}(T_1, T_0 < \infty) = \int \frac{\exp(\beta_0+\beta_z+\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{\beta}})}{1+\exp(\beta_0+\beta_z+\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{\beta}})} \frac{\exp(\beta_0+\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{\beta}})}{1+\exp(\beta_0+\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{\beta}})} dP(\tilde{\boldsymbol{X}})$, $\pi_1 = \mathbb{P}(T_1 = \infty) = \int \frac{1}{1+\exp(\beta_0+\beta_z+\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{\beta}})} dP(\tilde{\boldsymbol{X}})$, *and* $\pi_0 = \mathbb{P}(T_0 = \infty) = \int \frac{1}{1+\exp(\beta_0+\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{\beta}})} dP(\tilde{\boldsymbol{X}})$ *are the cured probabilities for the treatment and control arms.*

See proof in Appendix A.

The following Remarks reveal that the unified estimand connects and bridges the risk difference definition from the binary endpoint and the hazard ratio definition from the survival endpoint.

Remark 5. When $\beta_0 \to \infty$, the cured probabilities converge to 0 for both the treatment and control arms ($\lim_{\beta_0\to\infty} \pi_0 = \lim_{\beta_0\to\infty} \pi_1 = 0$), the model reduces to the proportional hazards model.

$$\Delta_M \to \Delta_{PH} = \tanh(-\frac{\gamma_z}{2}).$$

Remark 6. If there is no treatment effect for the non-cured population ($\gamma_z = 0$), $\Delta_M = \pi_1 - \pi_0$, which coincides with the definition of risk difference.

Theorem 4. *Under the promotion time cure model defined in (2.15), let ζ_z denote the log hazard ratio between the treatment and control arms. The unified estimand is a one-to-one transformation of the log hazard ratio ζ_z.*

$$\Delta_P = \frac{1-\exp(\zeta_z)}{1+\exp(\zeta_z)} = \tanh(-\frac{\zeta_z}{2}).$$

Theorem 4 is straightforward following Theorem 2 and thus the proof is omitted.

4. BAYESIAN INFERENCE UNDER CURE MODELS

4.1 Priors and Posteriors

4.1.1 Mixture Cure Model

For the mixture cure model, it has been shown in Theorem 2 of [3] that if we take an improper uniform prior for $\boldsymbol{\beta}$ [i.e., $\pi_0(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha) \propto \pi_0(\boldsymbol{\gamma}, \alpha)$], the posterior distribution

$$\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha \mid \mathcal{D}^{obs}) \propto L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha \mid \mathcal{D}^{obs})\pi_0(\boldsymbol{\gamma}, \alpha)$$

is always improper regardless of the propriety of $\pi_0(\boldsymbol{\gamma}, \alpha)$, i.e.,

$$\int_\Theta \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha \mid \mathcal{D}^{obs}) d\boldsymbol{\theta} = \infty.$$

Thus, a uniform improper prior $\pi_0(\boldsymbol{\theta}) \propto 1$ cannot guarantee convergence of MCMC sampling. Instead, weak independent normal priors will be put on each parameter in $\boldsymbol{\beta}$, that is, $\pi_0(\boldsymbol{\theta}) \sim \prod_j \phi(\beta_j; 0, \sigma_\beta^2)$, where $\phi(x; \mu_x, \sigma^2)$ denotes the probability density function of the normal distribution $N(\mu_x, \sigma^2)$. The posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha)$ given the observed data $\mathcal{D}^{obs}$ can be written as

$$\begin{aligned} &\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha \mid \mathcal{D}^{obs}) \\ &\propto \mathcal{L}_M^{obs}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha \mid \mathcal{D}^{obs})\pi_0(\boldsymbol{\beta}). \end{aligned} \tag{4.1}$$

4.1.2 Promotion Time Cure Model

Suppose that we consider a joint weak informative prior for the parameters under the promotion time cure model. Suppose $\pi_0(\boldsymbol{\zeta}, \alpha, \mu) \propto \pi_0(\alpha \mid \nu_0, \tau_0)\pi_0(\mu)$, where $\pi_0(\alpha \mid \nu_0, \tau_0)$ is a gamma prior for shape parameter α, and $\pi_0(\mu)$ is a week normal prior for μ. The posterior distribution of $(\boldsymbol{\zeta}, \alpha, \mu)$ given the observed data $\mathcal{D}^{obs}$ can be written as

$$\begin{aligned} &\pi(\boldsymbol{\zeta}, \alpha, \mu \mid \mathcal{D}^{obs}) \\ &\propto \mathcal{L}_P^{obs}(\boldsymbol{\zeta}, \alpha, \mu \mid \mathcal{D}^{obs})\pi_0(\alpha \mid \nu_0, \tau_0)\pi_0(\mu). \end{aligned} \tag{4.2}$$

4.2 Bayesian Hypothesis Testing

We consider a clinical trial that aims to demonstrate that a new treatment presents treatment benefits for the population. Each patient will be randomly allocated into one of the two arms: the treatment arm and the active control arm. We assume the randomization ratio is fixed as 1:1, and the total sample size is denoted by n. Progression-free survival (PFS) or Relapse-free survival (RFS) is the primary endpoint, and the survival time and event indicator (t_i, δ_i) will be collected for each patient.

The general hypothesis can be expressed as

$$H_1(g) : g(\boldsymbol{\theta}) > 0 \text{ v.s. } H_0(g) : g(\boldsymbol{\theta}) \leq 0 \tag{4.3}$$

or

$$H_1(g) : g(\boldsymbol{\theta}) < 0 \text{ v.s. } H_0(g) : g(\boldsymbol{\theta}) \geq 0, \tag{4.4}$$

where the treatment effect parameter we are interested in, $g(\boldsymbol{\theta})$, can take any functional form of model parameters $\boldsymbol{\theta}$. Also, constructing the functional form of $g(\boldsymbol{\theta})$ may include prognostic factors as well.

Under the mixture cure model, denote the vector of parameters as $\boldsymbol{\theta}_M = \{\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha\}$. The treatment effect parameter may include

(i) $g_1(\boldsymbol{\theta}_M) = \beta_z$ for testing the *conditional* treatment effect in the cure fraction,
(ii) $g_2(\boldsymbol{\theta}_M) = \gamma_z$ for testing the *conditional* treatment effect in the non-cured survival component, and
(iii) $g_3(\boldsymbol{\theta}_M) = \Delta_M$ for testing the *unconditional* treatment effect in the unified estimand defined for MCM in Section 3.

Under the promotion time cure model, let $\boldsymbol{\theta}_P = \{\boldsymbol{\zeta}, \alpha, \mu\}$ be the vector of parameters. The treatment effect parameter could take the following form:

(iv) $g_4(\boldsymbol{\theta}_P) = \zeta_z$ for testing the *conditional* treatment effect in log hazard ratio, and
(v) $g_5(\boldsymbol{\theta}_P) = \Delta_P$ for testing the *unconditional* treatment effect in the unified estimand defined for PTCM in Section 3.

To be noted that although Δ_M and Δ_P are defined based on different model assumptions, both of them represent the unified estimand Δ.

From Bayesian hypothesis testing point of view, the hypothesis $H_0(g)$ can be rejected and the data is in favor of $H_1(g)$ if the posterior predictive probability

$$\mathbb{P}(H_1(g) \mid \mathcal{D}^{(n)}) \geq \gamma^*, \tag{4.5}$$

where γ^* is a pre-specified threshold, which equals 0.95 by default, for rejecting $H_0(g)$. On the other hand, the null hypothesis, $H_0(g)$, cannot be rejected if $\mathbb{P}(H_1(g) \mid \mathcal{D}^{(n)}) < \gamma^*$.

4.3 Model Comparisons

In an ideal scenario, clinicians and researchers ought to predetermine the choice between the two types of cure models based on an in-depth understanding of the disease mechanism and the nature of the intervention. Nevertheless, when the cure model type has not been pre-specified, goodness-of-fit (GoF) can be evaluated to facilitate selection and comparison. By using the unified estimand proposed, the model selection and variable selection processes do not alter the definition of the estimand since it reflects the unconditional treatment effect.

In this section, we consider two Bayesian model comparison criteria, including the Deviance Information Criterion (DIC) [22] and the Logarithm of Pseudo-Marginal Likelihood (LPML) [14]. These two goodness-of-fit measures can help choose from the two types of cure model when it is needed.

The DIC is defined as

$$\mathrm{DIC} = \mathrm{Dev}(\overline{\boldsymbol{\theta}}) + 2p_D,$$

where deviance $\mathrm{Dev}(\boldsymbol{\theta}) = -2\log(\mathcal{L}^{obs}(\boldsymbol{\theta} \mid \mathcal{D}^{obs}))$ and $\mathcal{L}^{obs}(\boldsymbol{\theta} \mid \mathcal{D}^{obs})$ refers to either (2.9) or (2.17) depending on the model, $\mathrm{Dev}(\overline{\boldsymbol{\theta}})$ is the deviance evaluated at the posterior mean of $\boldsymbol{\theta}$, and $p_D = \overline{\mathrm{Dev}(\boldsymbol{\theta})} - \mathrm{Dev}(\overline{\boldsymbol{\theta}})$ represents the effective number of parameters. The DIC offers a measure of the trade-off between the fit of the model and its complexity, selecting models that provide a good fit without overfitting.

Besides DIC, the LPML is another commonly used criterion for Bayesian model comparison, which is a summary statistics based on CPO_i's and judges the predictive performance of different models based on the predictive distribution for each observation. The LPML is defined as

$$\mathrm{LPML} = \sum_i \log(\mathrm{CPO}_i). \tag{4.6}$$

Under the cure models, we define the conditional predictive ordinate (CPO) statistics for the i-th subject as

$$\begin{aligned}\mathrm{CPO}_i &= f(t_i \mid \boldsymbol{x}_i, \mathcal{D}^{obs(-i)})^{\delta_i} S(t_i \mid \boldsymbol{x}_i, \mathcal{D}^{obs(-i)})^{1-\delta_i} \\ &= \int f_i(t_i \mid \boldsymbol{x}_i, \boldsymbol{\theta})^{\delta_i} S_i(t_i \mid \boldsymbol{x}_i, \boldsymbol{\theta})^{1-\delta_i} \pi(\boldsymbol{\theta} \mid \mathcal{D}^{obs(-i)}) d\boldsymbol{\theta},\end{aligned}$$

where $\mathcal{D}^{obs(-i)}$ is the observed data $\mathcal{D}^{obs}$ with the i-th observation ignored, and $\pi(\boldsymbol{\theta} \mid \mathcal{D}^{(-i)})$ is the posterior given $\mathcal{D}^{obs(-i)}$ given in (4.1) or (4.2).

By [4],

$$\mathrm{CPO}_i = \left(\int \frac{1}{f_i(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{\delta_i} S_i(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{1-\delta_i}} \pi(\boldsymbol{\theta} \mid \mathcal{D}^{obs}) d\boldsymbol{\theta} \right)^{-1},$$

and it can be approximated via posterior computation after obtaining the MCMC samples $\{\boldsymbol{\theta}^{(l)}, l = 1, \ldots, L\}$ from the posterior distribution $\pi(\boldsymbol{\theta} \mid \mathcal{D}^{obs})$ by

$$\widehat{\mathrm{CPO}_i} = \left[\frac{1}{L} \sum_{l=1}^{L} \frac{1}{f_i(t_i|\boldsymbol{x}_i\, \boldsymbol{\theta}^{(l)})^{\delta_i} S_i(t_i|\boldsymbol{x}_i, \boldsymbol{\theta}^{(l)})^{1-\delta_i}} \right]^{-1}.$$

Plugging it in (4.6), we have

$$\widehat{\mathrm{LPML}} = \sum_i \log(\widehat{\mathrm{CPO}_i}).$$

4.4 Bayesian Computation

4.4.1 Mixture Cure Model

Under the mixture cure model, conditioning on the hidden state $\mathcal{Y}$, the cure rate component and the non-cured survival component can be separated, and both are log-concave to the set of parameters. The Gibbs sampler can be developed to draw MCMC samples from the posterior distribution by an accept-reject sampling scheme, which yields dependent samples.

The pseudo-algorithm for the updating scheme is written in Algorithm 1.

Algorithm 1 Gibbs sampler for the mixture cure model with normal initial priors.

Initialization: $\boldsymbol{\beta}^{(0)} \leftarrow \widehat{\boldsymbol{\beta}}$, $\boldsymbol{\gamma}^{(0)} \leftarrow \widehat{\boldsymbol{\gamma}}$, $\alpha^{(0)} \leftarrow \widehat{\alpha}$, $y_i^{(0)} \leftarrow 1 - \delta_i$.
while $l < L$ **do**
  Update $\boldsymbol{\beta}$: $\boldsymbol{\beta}^{(l)} \leftarrow q(\boldsymbol{\beta} \mid \mathcal{Y}^{(b-1)}) \propto \exp(\sum_{i=1}^n (1 - y_i^{(b-1)}) \boldsymbol{x}_i^\top \boldsymbol{\beta} - \log(1 + \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}))) \pi_0(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha)$;
  Update π_i by (2.6);
  Update $\boldsymbol{\gamma}, \alpha$: $q(\boldsymbol{\gamma}^{(l)}, \alpha^{(l)}) \leftarrow q(\boldsymbol{\gamma}, \alpha \mid \mathcal{Y}^{(b-1)}) \propto \alpha^{\sum_{y_i^{(l)}=0} \delta_i} \prod_{y_i^{(l)}=0} (t_i^{\delta_i(\alpha-1)} \exp(-t_i^\alpha \exp(\boldsymbol{x}_i^\top \boldsymbol{\gamma}^{(b-1)}))) \pi_0(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha)$.
  Update $\mathcal{Y}$:
  $y_i^{(l)} \leftarrow \min\{Bernoulli(\frac{\pi_i^{(l)}}{\pi_i^{(l)} + (1-\pi_i^{(l)}) S_i(t_i|\boldsymbol{x}_i, \boldsymbol{\gamma}^{(l)})}), 1 - \delta_i\}$;

After MCMC samples $\{\boldsymbol{\theta}_M^{(l)}, l = 1, \ldots, L\}$ are obtained, $\Delta_M^{(l)} = g_3(\boldsymbol{\theta}_M^{(l)})$ for the l-th iteration can be estimated by g-estimation as

$$\begin{aligned}\Delta_M^{(l)} &= g_3(\boldsymbol{\theta}_M^{(l)}) \\ &= \tanh(-\frac{\gamma_z^{(l)}}{2}) \frac{1}{n} \sum_{i=1}^n \left((1 - \pi_{i0}^{(l)})(1 - \pi_{i1}^{(l)}) \right) + \\ &\quad \frac{1}{n} \sum_{i=1}^n \left(\pi_{i1}^{(l)} - \pi_{i0}^{(l)} \right),\end{aligned} \tag{4.7}$$

where

$$\pi_{i1}^{(l)} = \frac{1}{1+\exp(\beta_0^{(l)} + \beta_z^{(l)} + \tilde{\boldsymbol{x}}_i^\top \tilde{\boldsymbol{\beta}}^{(l)})} \text{ and}$$
$$\pi_{i0}^{(l)} = \frac{1}{1+\exp(\beta_0^{(l)} + \tilde{\boldsymbol{x}}_i^\top \tilde{\boldsymbol{\beta}}^{(l)})} \tag{4.8}$$

are the l-th posterior predicted cured probabilities for patient i if the i-th patient is assigned to the treatment or control arms, correspondingly.

Then, the posterior mean $\overline{\Delta_M} = \frac{1}{L}\sum_{l=1}^{L} \Delta_M^{(l)}$ will be used as a point estimate of Δ_M from the Bayesian computation, and the posterior inference can be made based on $\{\Delta_M^{(l)}, l = 1, \ldots, L\}$.

4.4.2 Promotion Time Cure Model

After MCMC samples $\{\boldsymbol{\theta}_P^{(l)}, l = 1, \ldots, L\}$ are obtained, $\Delta_P^{(l)} = g_5(\boldsymbol{\theta}_P^{(l)})$ for the l-th iteration can be estimated by g-estimation as

$$\Delta_P^{(l)} = g_5(\boldsymbol{\theta}_P^{(l)}) = \tanh(-\frac{\zeta_z^{(l)}}{2}). \tag{4.9}$$

The posterior mean $\overline{\Delta_P} = \frac{1}{L}\sum_{l=1}^{L} \Delta_P^{(l)}$ will be used as a point estimate of Δ_P from the Bayesian computation, and the posterior inference can be made based on $\{\Delta_M^{(l)}, l = 1, \ldots, L\}$.

5. SIMULATION

In the simulation study, we simulate independent data from (2.8) for mixture cure models or (2.15) for promotion time cure models with $\boldsymbol{\theta}^{true}$ based on a 1:1 randomization ratio with arm sizes $n_1 = n_0 = 200$. The covariates $\tilde{\boldsymbol{x}}$, including sex and gender, for each patient are randomly drawn from the melanoma cancer E1684 dataset. More details about the E1684 dataset are introduced in Section 6. The time-to-event data is generated with independent censoring, but no maximum follow-up period is set.

Our main focus is on comparing the performance of the Bayesian inference with different choices of the five functional forms of treatment effect parameters $g(\boldsymbol{\theta})$ mentioned in Section 4.2. In addition, we evaluate and compare the performance when the model is either correctly specified or misspecified between the two types of cure models. Based on the posterior samples obtained from MCMC, the rejection probability (RP), root mean square error (RMSE), and coverage probability (CP) serve as Frequentist evaluations of the Bayesian inference for those parameters based on B replicates. When the model is correctly specified, we assess the RP, RMSE, and CP of the treatment effect parameter $g(\boldsymbol{\theta})$ by

$$RP(g(\boldsymbol{\theta})) = \frac{1}{B}\sum_{b=1}^{B}\left[1\{0 \in (g(\boldsymbol{\theta})^{LL(b)}, g(\boldsymbol{\theta})^{UL(b)})\}\right],$$

$$RMSE(g(\boldsymbol{\theta})) = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\overline{g(\boldsymbol{\theta})}^{(b)} - g(\boldsymbol{\theta}^{true})\right)^2},$$

and

$$CP(g(\boldsymbol{\theta})) = \frac{1}{B}\sum_{b=1}^{B}\left[1\{g(\boldsymbol{\theta}^{true}) \in (g(\boldsymbol{\theta})^{LL(b)}, g(\boldsymbol{\theta})^{UL(b)})\}\right],$$

where $(g(\boldsymbol{\theta})^{LL(b)}, g(\boldsymbol{\theta})^{UL(b)})$ is the 95% highest posterior density (HPD) interval of $g(\boldsymbol{\theta})$ from the b-th replicate. However, when the type of cure model is misspecified, we assess the RP, RMSE, and CP of the unified estimand Δ by

$$RP(\Delta_{mis}) = \frac{1}{B}\sum_{b=1}^{B}\left[1\{0 \in (\Delta_{mis}^{LL(b)}, \Delta_{mis}^{UL(b)})\}\right],$$

$$RMSE(\Delta_{mis}) = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\overline{\Delta_{mis}}^{(b)} - \Delta^{true}\right)^2},$$

and

$$CP(\Delta_{mis}) = \frac{1}{B}\sum_{b=1}^{B}\left[1\{\Delta^{true} \in (\Delta_{mis}^{LL(b)}, \Delta_{mis}^{UL(b)})\}\right],$$

where $\Delta^{true} = g(\boldsymbol{\theta}^{true})$ is the associated with the true model, $\overline{\Delta_{mis}}^{(b)}$ is the posterior mean and $(\Delta_{mis}^{LL(b)}, \Delta_{mis}^{UL(b)})$ is the 95% highest posterior density (HPD) interval of Δ based on the misspecified model from the b-th replicate.

To fit the simulated data by the mixture cure model, we apply independent normal initial priors for each parameter in $\boldsymbol{\beta}$ to ensure identifiability. The variance of the initial prior is set to $\sigma_\beta^2 = 100$ for each parameter of $\boldsymbol{\beta}$. We set the initial prior for the shape parameter α to $\pi^{(f)}(\alpha) \propto 1$. To sum up, the initial prior is set as $\pi^{(f)}(\boldsymbol{\theta}_M) = \prod_{j=1} \phi(\beta_j; 0, 10^2)$. For the initial priors of the promotion time cure model, we set $\pi^{(f)}(\zeta_j) \sim \mathcal{N}(0, 100^2)$ for each dimension in $\boldsymbol{\zeta}$, consider the shape parameter α as unknown, and set the prior for μ and α to be flat $\pi^{(f)}(\mu, \alpha) \propto 1$. Therefore, for the promotion time cure model, the initial prior is set as $\pi^{(f)}(\boldsymbol{\theta}_P) = \prod_{j=1} \phi(\zeta_j; 0, 100^2)$.

All Bayesian simulation studies are based on Markov Chain Monte Carlo samples of size $L = 1000$, with additional 1000 burn-in iterations unless otherwise noted. We set $\gamma^* = 0.95$ for making the Bayesian decision. $B = 30,000$ replications are used for evaluating the performance under null hypothesis at 5% level of significance and $B = 5,000$ replications are used for evaluating the performance under alternative hypothesis.

5.1 Simulation Results

The MCMC samples demonstrate adequate convergence and mixing across all examined scenarios, which supports the posterior propriety and model identifiability. The MCMC trace plot and mixing diagnoses are thus omitted.

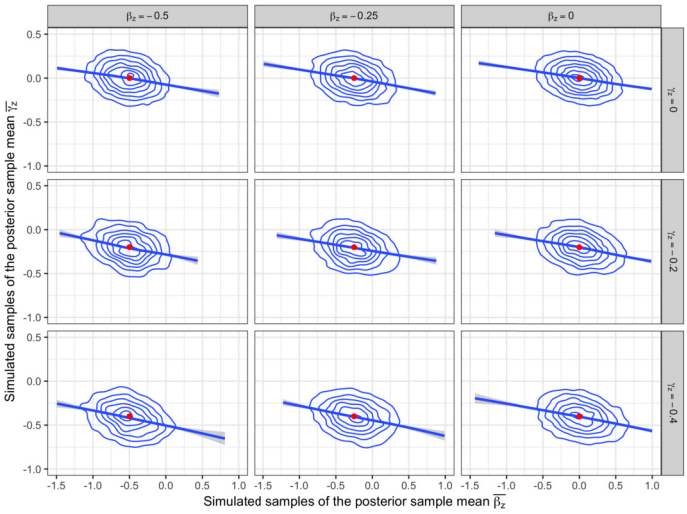

Figure 5.1: The contour plots of the posterior sample means $\overline{\beta_z}$ and $\overline{\gamma_z}$ with different β_z, γ_z.

5.1.1 The Mixture Cure Model

For the true model, we choose β_z from $\{0, -0.25, -0.5\}$, which correspond to odds ratios of $\{1, 0.78, 0.61\}$ for the cured probability, and γ_z from $\{0, -0.2, -0.4\}$, which correspond to hazard ratios of $\{1, 0.82, 0.67\}$ for the non-cured population. We specify the shape of the Weibull distribution as 1 and assume an independent censoring rate of 0.2. The other hyperparameters are set to be $\beta_0 = 1$, $\tilde{\boldsymbol{\beta}} = (0.3, -0.5)^\top$, $\gamma_0 = 0$, and $\tilde{\boldsymbol{\gamma}} = (0.2, -0.4)^\top$ such that the event rates are ranged between 40% and 60%.

Figure 5.1 presents contour plots of the posterior samples means $\overline{\beta_z}$ and $\overline{\gamma_z}$ obtained from the mixture cure model. Different subplots correspond to different β_z and γ_z in the true model, with the red point indicating the true values of (β_z, γ_z). The plots illustrate that the posterior sample means converge and the mode of the posterior sample means is close to the truth of β_z and γ_z. Remarkably, the negative correlations stated in Remark 2 are evident in all scenarios.

The results of $RP(g(\boldsymbol{\theta}))$ under the mixture cure model are shown in Table 5.1. Each row represents different true values for β_z, γ_z, and their resulting Δ_M. When $\beta_z = \gamma_z = 0$, $RP(g(\boldsymbol{\theta}))$ characterizes the probability of rejecting $H_0(g)$ under null hypothesis (false rejection) for the treatment effect parameters $g(\boldsymbol{\theta})$. As shown in the first row of Table 5.1, $RP(\beta_z)$, $RP(\gamma_z)$, and $RP(\Delta_M)$ are right around the pre-specified level of significance 0.05. When the data are incorrectly fitted by the promotion time cure model, the rejection probability $RP(\Delta_P) = 0.054$ is slightly over the level of 0.05. When β_z and γ_z become larger in absolute value, $RP(g(\boldsymbol{\theta}))$ characterizes the probability of favoring $H_1(g)$ under alternative hypothesis (true rejection) for the treatment effect parameters $g(\boldsymbol{\theta})$. When the investigational treatment shows efficacy for the cure fraction only (row 4 and 7 in Table 5.1), the rejection probability of the unified estimand

Table 5.1. Results of $RP(g(\boldsymbol{\theta}))$ under MCM.

Truth of $\boldsymbol{\theta}_M$			Fitted model:	MCM			PTCM
β_z	γ_z	Δ_M	$g(\boldsymbol{\theta})$:	β_z	γ_z	Δ_M	Δ_P
0	0	0	$RP(g(\boldsymbol{\theta}))$:	0.053	0.052	0.051	0.054
0	−0.2	0.052		0.049	0.373	0.234	0.212
0	−0.4	0.103		0.049	0.812	0.546	0.481
-0.25	0	0.050		0.235	0.054	0.214	0.230
-0.25	−0.2	0.098		0.204	0.372	0.499	0.498
-0.25	−0.4	0.146		0.191	0.817	0.830	0.751
-0.5	0	0.103		0.602	0.052	0.547	0.565
-0.5	−0.2	0.148		0.561	0.376	0.819	0.799
-0.5	−0.4	0.192		0.552	0.799	0.963	0.937

Δ_M, $RP(\Delta)$, is slightly smaller than $RP(\beta_z)$. When the treatment shows efficacy on the non-cured survival only (row 2 and 3 in Table 5.1), $RP(\Delta)$ is about 60%–70% of $RP(\gamma_z)$. When the investigational treatment shows efficacy for both cure fraction and non-cured survival (row 5,6,8, and 9 in Table 5.1), the rejection probability of the unified estimand Δ_M, $RP(\Delta)$, is greater than $RP(\beta_z)$ or $RP(\gamma_z)$ when looking at β_z and γ_z alone. Also, when the model is misspecified, the $RP(\Delta)$ remains comparable with the correctly specified model in all simulation circumstances.

Table 5.2 displays the results of root mean square error (RMSE) and coverage probability (CP). Within the Bayesian framework, the initial prior with a variance of 100 performs well recovering the truth since the RMSEs of all treatment effect parameters are relatively low and the coverage probabilities are close to 0.95. The coverage probabilities are slightly lower than 0.95 when an informative prior (shrinkage prior) is implemented when constructing the initial prior. The estimation is more accurate for the unified estimand Δ than looking at the conditional treatment effect parameters $\{\beta_z, \gamma_z\}$ since the RMSE of Δ is close to zero for each and every simulation scenario. For the misspecified model fittings, the RMSEs are slightly larger than the correctly-specified model, and the coverage probabilities are slightly smaller. The difference appears to increase as β_z and γ_z deviate from 0, but the RMSE and the CP for the misspecified model are pretty robust overall.

5.1.2 The Promotion Time Cure Model

For the true model, we choose ζ_z from $\{0, -0.25, -0.5\}$, corresponding to hazard ratios of $\{1, 0.78, 0.61\}$ for the conditional treatment effect. We specify the shape of the Weibull survival time to be 1 with an independent censoring rate of 0.1. The other hyperparameters are set as $\zeta_0 = 1$ and $\tilde{\boldsymbol{\zeta}} = (0.3, -0.5)^\top$.

The results of $RP(g(\boldsymbol{\theta}))$ under the promotion time cure model are shown in Table 5.3. When ζ_z and its resulting Δ_P are 0, $RP(g(\boldsymbol{\theta}))$ represents the probability of rejecting $H_0(g)$ under null hypothesis (false rejection) for $g(\boldsymbol{\theta})$. As shown in the first row of Table 5.3, $RP(\zeta_z) = 0.053$ and $RP(\Delta_P) = 0.053$. When the model is misspecified as

Table 5.2. Results of Bayesian $RMSE(g(\boldsymbol{\theta}))$ and $CP(g(\boldsymbol{\theta}))$ under MCM.

Truth of $\boldsymbol{\theta}_P$			Fitted Models (f):	$RMSE(g(\boldsymbol{\theta}))$				$CP(g(\boldsymbol{\theta}))$			
				MCM			PTCM	MCM			PTCM
β_z	γ_z	Δ_M	$g(\boldsymbol{\theta})$:	β_z	γ_z	Δ_M	Δ_P	β_z	γ_z	Δ_M	Δ_P
0	0	0		0.002	0.002	0.0001	0.0002	0.944	0.941	0.940	0.937
0	−0.2	0.052		0.021	0.005	0.001	0.001	0.944	0.942	0.945	0.937
0	−0.4	0.103		0.026	0.008	0.001	0.002	0.945	0.941	0.938	0.935
-0.25	0	0.050		0.007	0.005	0.0002	0.009	0.940	0.940	0.943	0.933
-0.25	−0.2	0.098		0.014	0.006	0.002	0.011	0.942	0.943	0.945	0.931
-0.25	−0.4	0.146		0.005	0.019	0.005	0.012	0.941	0.935	0.941	0.931
-0.5	0	0.103		0.026	0.004	0.001	0.019	0.943	0.942	0.942	0.920
-0.5	−0.2	0.148		0.017	0.014	0.003	0.021	0.943	0.936	0.940	0.922
-0.5	−0.4	0.192		0.026	0.014	0.006	0.023	0.940	0.942	0.940	0.917

Table 5.3. Results of $RP(g(\boldsymbol{\theta}))$ under PTCM.

Truth of $\boldsymbol{\theta}_P$		Fitted Models (f):	PTCM		MCM		
ζ_z	Δ_P	$g(\boldsymbol{\theta})$:	ζ_z	Δ_P	β_z	γ_z	Δ_M
0	0	$RP(g(\boldsymbol{\theta}))$:	0.053	0.053	0.037	0.068	0.058
-0.25	0.124		0.612	0.612	0.070	0.583	0.636
-0.5	0.24		0.981	0.981	0.132	0.899	0.980

Table 5.4. Results of Bayesian $RMSE(g(\boldsymbol{\theta}))$ and $CP(g(\boldsymbol{\theta}))$ under PTCM.

Truth of $\boldsymbol{\theta}_P$		Fitted Models:	$RMSE(g(\boldsymbol{\theta}))$			$CP(g(\boldsymbol{\theta}))$		
			PTCM		MCM	PTCM		MCM
ζ_z	Δ_P	$g(\boldsymbol{\theta})$:	ζ_z	Δ_P	Δ_M	ζ_z	Δ_P	Δ_M
0	0		0.001	0.0005	0.001	0.942	0.941	0.927
-0.25	0.124		0.005	0.001	0.004	0.936	0.934	0.924
-0.5	0.24		0.007	0.001	0.005	0.938	0.935	0.920

the mixture cure model, $RP(\beta_z)$ is deflated to 0.037, and $RP(\gamma_z)$ is inflated to 0.068. Hence, making inference based on β_z and γ_z will not maintain the rejection probability at the proper level. However, the $RP(\Delta_M)$ is less affected by model misspecification, which is 0.058.

When the assumed treatment effect becomes larger as ζ_z decreases from 0, $RP(\Delta_P)$ is similar to $RP(\zeta_z)$ considering the one-to-one relationship under promotion time cure model shown in theorem 4. If the model is misspecified, among the examined scenarios, the rejection probability $RP(\Delta_M)$ is higher than $RP(\beta_z)$ or $RP(\gamma_z)$. In addition, $RP(\Delta_M)$ is comparable to the rejection probability $RP(\Delta_P)$ from the correctly specified model.

Table 5.4 displays the results of root mean square error (RMSE) and coverage probability (CP). Within the Bayesian framework, the models with a variance of 100^2 in the initial prior converge pretty well and could recover the truth as the RMSEs of treatment effect parameters ζ_z and Δ_P are all rather low and the coverage probabilities are close to 0.95. The coverage probabilities with model misspecification are slightly lower, and RMSEs are slightly larger than those fitted by the correct model, but the misspecified model remains comparable with the correctly specified model.

6. REAL DATA EXAMPLE

We consider data from the Eastern Cooperative Oncology Group's phase III melanoma clinical trial, E1684 [15], involving three treatment arms, namely, high-dose interferon, low-dose interferon, or observation. The results of the E1684 clinical trial suggested that interferon has a significant impact on relapse-free and overall survival, which led to the approval of this regimen by the U.S. Food and Drug Administration as standard adjuvant therapy for high-risk melanoma patients. For our analysis, we only consider the high-dose interferon (treatment) and observation (control).

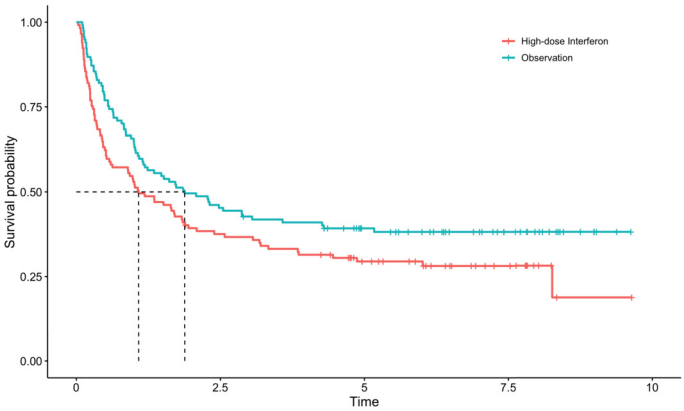

Figure 6.1: Kaplan-Meier plot for E1684.

The E1684 dataset includes four predictors: treatment (x_1, coded as 2 for high-dose interferon and 1 for observation), type of primary tumor (x_2, coded as 2 for nodular and 1 for other), age (x_3, coded in years), and gender (x_4, coded as 2 for female and 1 for male). We include treatment, age, and gender as predictors in our real data example.

The outcome variable is relapse-free survival (RFS), which is defined as the time from randomization until the progression of the tumor or the occurrence of mortality, whichever comes first. The RFS time, t_i in years, is continuous and subject to right censoring. δ_i denotes the censoring indicator, which equals one if the i-th subject relapsed and 0 otherwise.

After eliminating the incomplete data from the E1684 dataset, we have a total of $n = 284$ patients, with $n_1 = 144$ in the high-dose interferon arm and $n_0 = 140$ in the observation arm. The Kaplan-Meier curves for the treatment and placebo arm of E1684 are shown in Figure 6.1. The Kaplan-Meier curves demonstrate plateau patterns of the cure models after a sufficiently long follow-up period. The median relapse-free survival time (95% confidence interval) for the high-dose interferon and the observation arms are 1.72 $(1.09, 3.02)$ and 0.98 $(0.52, 1.70)$, respectively. The median overall survival (OS) time (95% confidence interval) for the high-dose interferon and the observation arms are 3.82 $(2.56, \infty)$ and 2.67 $(1.83, 4.24)$, respectively.

We carry out a comprehensive analysis for the relapse-free survival (RFS) of the E1684 dataset using different models, including the Cox proportional hazards model, restricted mean survival time comparison, mixture cure model, and promotion time cure model. The effect of age and gender are adjusted in all comparisons, and the unified estimand are analyzed and compared with individual model parameters under each model. The results are summarised in Table 6.1.

Under the Cox proportional hazards model, the conditional treatment effect of log hazards ratio β_z between the treatment and control arms shows significance ($\widehat{\beta_z} = -0.360$, 95% confidence interval: $(-0.642, -0.079)$), resulting in a significant treatment effect of the unified estimand Δ_{PH} ($\widehat{\Delta_{PH}} = 0.178$, 95% confidence interval: $(0.039, 0.310)$). The p-values associated with testing β_z or Δ_{PH} are 0.006. For the RMST comparison, the results are sensitive to the choice of the truncation time t^*, and could lead to entirely different conclusion. When the truncation time is set to be the last observed event (by default), the difference in restricted mean survival time (RMST) between the treatment and control groups $\widehat{\Delta_{RMST}}(9.63) = -0.022$ with a 95% confidence interval $(-1.156, 1.112)$ and p-value 0.970. When the truncation time is set to be median RFS (1.24 months), $\widehat{\Delta_{RMST}}(1.24) = 0.148$ with a 95% confidence interval $(0.043, 0.253)$ and p-value 0.006.

Under the mixture cure model, when looking at the conditional treatment effect parameters, the conditional treatment effect on the cured probability β_z is significant (posterior mean: -0.596, 95% credible interval: $(-1.169, -0.067)$, posterior probability $\mathbb{P}(\beta_z < 0 \mid \mathcal{D}^{obs}) = 0.986$), but the conditional treatment effect on non-cured population γ_z is not (posterior mean: -0.103, 95% credible interval: $(-0.409, 0.221)$, posterior probability $\mathbb{P}(\gamma_z < 0 \mid \mathcal{D}^{obs}) = 0.742$)). However, the unconditional treatment effect on the unified estimand Δ_M is significant (posterior mean: 0.114, 95% credible interval: $(0.005, 0.231)$, posterior probability $\mathbb{P}(\Delta_M > 0 \mid \mathcal{D}^{obs}) = 0.978$).

Under the promotion time cure model, the treatment shows significant efficacy in the conditional treatment effect

Table 6.1. A comparison of E1684 results under different models.

Fitted Models:[1]	Cox PH		RMST		MCM			PTCM	
Parameter $g(\boldsymbol{\theta})$:	γ_z	Δ_{PH}	$\Delta_{RMST}(9.63)$	$\Delta_{RMST}(1.24)$	β_z	γ_z	Δ_M	ζ_z	Δ_P
Estimation[1]	-0.360	0.178	-0.022	0.148	-0.596	-0.103	0.114	-0.376	0.185
95% lower CI[1]	-0.642	0.039	-1.156	0.043	-1.169	-0.409	0.005	-0.669	0.055
95% upper CI[1]	-0.079	0.310	1.112	0.253	-0.067	0.221	0.231	-0.109	0.322
p[1]	0.006	0.006	0.970	0.006	0.014	0.258	0.022	0.002	0.002

The Cox PH model and Restricted Mean Survival Time (RMST) comparison are performed under the Frequentist framework, while the MCM and PTCM are fitted by Bayesian procedures. The "parameter estimation" refers to MLE or posterior mean, the "CI" refers to 95% two-sided confidence intervals or 95% HPD intervals, and p refers to the two-sided p-value or posterior predictive probability of $\mathbb{P}(H_0(g) \mid \mathcal{D}^{obs})$ depending on the models.

parameter ζ_z (posterior mean: -0.376, 95% credible interval: $(-0.669, -0.109)$, posterior probability $\mathbb{P}(\zeta_z < 0 \mid \mathcal{D}^{obs}) = 0.998$), resulting in significance in the unified estimand Δ_P (posterior mean: 0.185, 95% credible interval: $(0.055, 0.322)$, posterior probability $\mathbb{P}(\Delta_P > 0 \mid \mathcal{D}^{obs}) = 0.998$).

The two types of cure models are rigorously compared using the goodness-of-fit (GoF) criteria introduced in Section 4.3. For the MCM, the Deviance Information Criterion (DIC) stands at 772.01 with $p_d = 8.81$, whereas for the PTCM, it's 765.95 with $p_d = 6.19$. Additionally, the Pseudo-Marginal Likelihood (LPML) values for MCM and PTCM are -386.72 and -382.97, respectively. In summary, these metrics suggest that the promotion time cure model exhibits a better fit and displays greater predictive power, as evidenced by its reduced DIC and enhanced LPML values.

For the mixture cure model, the high-dose interferon presents significance in β_z but insignificance in γ_z, which confuses the overall efficacy of high-dose interferon and making a final decision. However, the results of the unified estimand are easier to interpret, and it allows direct comparisons across different models. Other than the inconsistent inference from RMST comparisons, the other three survival models lead to similar values for the estimators of Δ and reach the same conclusions. The similarity in the values of the estimators from different models reveals the robustness of the unified estimand against model selections between different types of cure models in capturing the overall unconditional treatment effect. Therefore, the proposed estimand provides a unified way to compare the efficacy on treatment and aids in making informed clinical decisions.

7. DISCUSSIONS

In cancer research, time-to-event endpoints, such as progression-free survival (PFS) and overall survival (OS), are critical in evaluating the efficacy of cancer treatments. However, this process is often time-consuming and costly, necessitating advanced modeling and inference strategies about the treatment effect in drug development [25]. In recent years, cure models have gained increasing interest due to their ability to account for a subset of patients considered "cured," who remain event-free after a certain follow-up period. These models provide more flexible and precise treatment outcome modeling and enhance decision-making about the treatment effect in clinical trials.

In this article, we have summarised two types of cure models and have proposed a unified estimand $\Delta = \mathbb{E}[sgn(T_1 - T_0)]$ for the survival models with cure fraction, where T_0 and T_1 denote the potential outcome when assigning to the control group and treatment group, and $sgn()$ denotes the sign function. The proposed unified estimand measures unconditional treatment effect by examining whether the treatment extends survival for patients. The relationships between the unified estimand Δ and model parameters have been established explicitly for the proportional hazards model, the mixture cure model, and the promotion time cure model, making comparisons across different models more straightforward. Based on the recent hot topic about the "Estimand" and "Covariate Adjustment" from FDA's guidances, including covariates that are prognostic in the regression models could potentially increase the power of RCTs, or it could make the results more interpretable in non-randomized trials. However, it is well noted that the estimand from the adjusted analysis about the conditional treatment effect is not always the same estimand for the unconditional treatment effect via direct comparison. The proposed unified estimand Δ is invariant to model and variable selections, and it can be applied to different types of cure models either without covariates or in regression forms adjusting for the prognostic covariates, without altering the definition of treatment effect. Therefore, when the unified estimand is utilized under the "Estimand" framework, the inference based on different models with or without covariate adjustment can serve as sensitivity analyses towards the same unified estimand instead of supplementary analyses with an altered definition of treatment effect.

Compared with $\Delta_{RMST}(t^*)$, which is highly sensitive to the choice of truncation time t^* and the tail of the Kaplan-Meier curve, the proposed unified estimand Δ reflects an overall difference in survival profiles, providing a clear and robust basis for determining treatment efficacy. In addition, the value of $\Delta_{RMST}(t^*)$ alone does not clearly indicate the effect size without comparing it to the RMST of the control arm. However, the unified estimand Δ directly measures the size of the treatment effect by the proportion of patients who experience extended survival on treatment.

In addition to proposing the unified estimand, we describe a general Bayesian inference procedure for the cure models. The Bayesian computation for the cure models are developed via Gibbs sampling for Bayesian inference, and model comparisons among different types of survival models with cure fraction are made through DIC and LPML.

Based on the simulation study, the Bayesian inference of the unified estimand Δ maintains the probabilities of rejecting $H_0(g)$ under null hypothesis at the desired level for both cure models. The unified estimand Δ demonstrates larger probabilities of favoring $H_1(g)$ when treatment benefits both cure fraction and non-cured survival.

Also, the unified estimand Δ is robust against model misspecification between the two types of cure models. When data are generated from one of the cure models but fitted by the other one, the inference of the unified estimand Δ always leads to reasonable probability of rejecting $H_0(g)$ under null, probability of favoring $H_1(g)$ under alternative hypothesis, root mean square error, and coverage probability.

In the current simulation scenarios, σ_β^2 is chosen as 10^2 for MCM. When σ_β^2 decreases from ∞ to 1, the probability of rejecting $H_0(g)$ will be expected to decrease generally, and the RMSEs of the treatment effect estimators will decrease and then increase, which is similar to the trend from the

Ridge regressions or using Bayesian shrinkage priors. Thus, an appropriate variance for the initial prior should be specified under the mixture cure model. If σ_β^2 is too large, the prior will be less informative, and the parameters will suffer from a lack of identifiability, leading to unreliable inference. If σ_β^2 is too small, the prior imposes too much information on the model, which leads to a biased estimator. A trade-off choice of σ_β^2 will balance model identifiability and estimation unbiasedness. In the end, we conduct an in-depth analysis of ECOG's melanoma cancer data E1684 using the unified estimand Δ under different models, including the proportional hazards model, mixture cure model, and promotion time cure model.

All three models lead to similar results and consistent conclusions, which demonstrates Δ's consistency compared to the difference in RMST.

Although this article focuses on estimation and Bayesian inference, the proposed estimand and the Bayesian inference framework are also useful in the Bayesian clinical trial designs under survival models with cure fraction. However, it is crucial to note that the simulation under the null hypothesis in this article corresponds to a sharp null hypothesis, assuming all individual treatment effect parameters are zero. However, to develop a rigorous Bayesian clinical trial design applying the proposed estimand, the null hypothesis should be relaxed to only restrict the expectation form of the estimand Δ to be 0. Such a relaxation would impact the size of the hypothesis testing and the sample size determination. Consequently, there is a need for rigorous evaluations of the Bayesian type I error and Bayesian power, noting an important direction for further research.

All computations and implementations detailed in this paper are developed in the R environment. The corresponding code for the proposed methodologies is available at https://github.com/hongfei-li/UniCure.

APPENDIX A. PROOFS OF THEOREMS

Proof of Theorem 1. The log observed-data likelihood can be written as

$$\ell_M^{obs}(\boldsymbol{\theta}_M \mid \mathcal{D}^{obs}) = \sum_{\delta_i=1} (\log(1-\pi_i) + \log(f_i^{nc}(t_i \mid \boldsymbol{x}_i))) + \sum_{\delta_i=0} \log\left(\pi_i + (1-\pi_i) S_i^{nc}(t_i \mid \boldsymbol{x}_i)\right).$$

The (β_z, γ_z) entry of the observed Fisher information matrix is given by

$$\begin{aligned}
&\mathcal{I}(\boldsymbol{\theta}_M)_{(\beta_z,\gamma_z)} \\
&= -\frac{\partial^2}{\partial\beta_z\partial\gamma_z}\ell_M^{obs}(\boldsymbol{\theta}_M \mid \mathcal{D}^{obs}) \\
&= -\sum_{\delta_i=0} \frac{\partial^2}{\partial\beta_z\partial\gamma_z}\log(\pi_i + (1-\pi_i)S_i^{nc}(t_i \mid \boldsymbol{x}_i)) \\
&= \sum_{\delta_i=0} \frac{1}{(\pi_i + (1-\pi_i)S_i^{nc}(t_i \mid \boldsymbol{x}_i))^2}\frac{\partial\pi_i}{\partial\beta_z}\frac{\partial S_i^{nc}(t_i \mid \boldsymbol{x}_i)}{\partial\gamma_z} \\
&> 0.
\end{aligned}$$

□

Proof of Theorem 2. Suppose $S_{pop,0}(t)$ and $S_{pop,1}(t)$ are the proper survival functions with $\lim_{t\to\infty} S_{pop,0}(t) = \lim_{t\to\infty} S_{pop,1}(t) = 0$. Then,

$$\begin{aligned}
&\int_0^\infty S_{pop,1}(t) f_{pop,0}(t)dt + \int_0^\infty S_{pop,0}(t) f_{pop,1}(t)dt \\
&= -S_{pop,0}(t)S_{pop,1}(t)\Big|_0^\infty = 1, \text{ and}
\end{aligned} \tag{1.1}$$

$$\begin{aligned}
&\int_0^\infty S_{pop,1}(t) f_{pop,0}(t)dt - \int_0^\infty S_{pop,0}(t) f_{pop,1}(t)dt \\
&= \Delta_{PH}.
\end{aligned} \tag{1.2}$$

Since $h_{pop,1}(t) = \exp(\gamma_z)h_{pop,0}(t)$,

$$\begin{aligned}
&\int_0^\infty S_{pop,0}(t) f_{pop,1}(t)dt \\
&= \int_0^\infty S_{pop,0}(t)S_{pop,1}(t)h_{pop,1}(t)dt \\
&= \exp(\gamma_z)\int_0^\infty S_{pop,0}(t)S_{pop,1}(t)h_{pop,0}(t)dt \\
&= \exp(\gamma_z)\int_0^\infty S_{pop,1}(t) f_{pop,0}(t)dt.
\end{aligned} \tag{1.3}$$

Solving the matrix equations of (1.1)–(1.3) gives us

$$\Delta_{PH} = \frac{1-\exp(\gamma_z)}{1+\exp(\gamma_z)} = \tanh(-\frac{\gamma_z}{2}).$$

□

Proof of Theorem 3.

$$\begin{aligned}
&\Delta_M = \mathbb{E}_{\tilde{\boldsymbol{X}}}[\mathbb{E}[sgn(T_1 - T_0) \mid \tilde{\boldsymbol{X}}]] \\
&= \mathbb{E}_{\tilde{\boldsymbol{X}}}[\mathbb{P}(T_1, T_0 < \infty \mid \tilde{\boldsymbol{X}})\mathbb{E}[sgn(T_1 - T_0) \mid T_1, T_0 < \infty \mid \tilde{\boldsymbol{X}}] + \\
&\quad \mathbb{P}(T_1 = \infty, T_0 < \infty \mid \tilde{\boldsymbol{X}}) - \mathbb{P}(T_0 = \infty, T_1 < \infty \mid \tilde{\boldsymbol{X}})] \\
&= \mathbb{E}_{\tilde{\boldsymbol{X}}}[\mathbb{P}(T_1 < \infty \mid \tilde{\boldsymbol{X}})\mathbb{P}(T_0 < \infty \mid \tilde{\boldsymbol{X}})\tanh(-\frac{\gamma_z}{2}) + \\
&\quad \mathbb{P}(T_1 = \infty \mid \tilde{\boldsymbol{X}}) - \mathbb{P}(T_0 = \infty \mid \tilde{\boldsymbol{X}})] \\
&= \int \frac{\exp(\beta_0 + \beta_z + \tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{\beta}})}{1+\exp(\beta_0+\beta_z+\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{\beta}})}\frac{\exp(\beta_0 + \tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{\beta}})}{1+\exp(\beta_0+\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{\beta}})}dP(\tilde{\boldsymbol{X}}) \times \\
&\quad \tanh(-\frac{\gamma_z}{2}) + \int \frac{1}{1+\exp(\beta_0+\beta_z+\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{\beta}})}dP(\tilde{\boldsymbol{X}}) - \\
&\quad \int \frac{1}{1+\exp(\beta_0+\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{\beta}})}dP(\tilde{\boldsymbol{X}})
\end{aligned}$$

□

ACKNOWLEDGEMENTS

The authors would like to thank the Section Editors, the Associate Editor, and a reviewer for their very helpful com-

ments and suggestions, which have greatly improved our paper.

Accepted 20 July 2024

REFERENCES

[1] Berkson, J. and Gage, R. P. (1952). Survival Curve for Cancer Patients. *Journal of American Statistical Association* **47**(259) 501–515.

[2] Breslow, N. E., Edler, L. and Berger, J. (1984). A Two-Sample Censored-Data Rank Test for Acceleration. *Biometrics* **40**(4) 1049–1062. https://doi.org/10.2307/2531155.

[3] Chen, M. -H., Ibrahim, J. G. and Sinha, D. (1999). A New Bayesian Model for Survival Data with a Surviving Fraction. *Journal of the American Statistical Association* **94**(447) 909–919. https://doi.org/10.1080/01621459.1999.10474196. MR1723307

[4] Chen, M. -H., Shao, Q. -M. and Ibrahim, J. G. (2000). Other Topics. In *Monte Carlo Methods in Bayesian Computation* 307–355. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4612-1276-8_10.

[5] Chen, Q., Zeng, D., Ibrahim, J. G., Chen, M. H., Pan, Z. and Xue, X. (2015). Quantifying the average of the time-varying hazard ratio via a class of transformations. *Lifetime Data Analysis* **21**(2) 259–279. https://doi.org/10.1007/s10985-014-9301-0. MR3324230

[6] Cox, T. F. and Czanner, G. (2016). A practical divergence measure for survival distributions that can be estimated from Kaplan-Meier curves. *Statistics in Medicine* **35**(14) 2406–2421. https://doi.org/10.1002/sim.6868. MR3513521

[7] Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38**(4) 1041–1046. https://doi.org/10.2307/2529885.

[8] Farewell, V. T. (1986). Mixture models in survival analysis: Are they worth the risk? *Canadian Journal of Statistics* **14**(3) 257–262. https://doi.org/10.2307/3314804. MR0859638

[9] Fleming, T. R. and Harrington, D. P. (1981). A class of hypothesis tests for one and two sample censored survival data. *Communications in Statistics - Theory and Methods* **10**(8) 763–794. https://doi.org/10.1080/03610928108828073.

[10] Food and Drug Administration (2021). *E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials*. https://www.federalregister.gov/documents/2021/05/12/2021-10066/e9r1-statistical-principles-for-clinical-trials-addendum-estimands-and-sensitivity-analysis-in.

[11] Food and Drug Administration (2023). *Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products*. https://www.federalregister.gov/documents/2023/05/26/2023-11263/adjusting-for-covariates-in-randomized-clinical-trials-for-drugs-and-biological-products-guidance.

[12] Gray, R. J. and Tsiatis, A. A. (1989). A linear rank test for use when the main interest is in differences in cure rates. *Biometrics* **45**(3) 899–904. https://doi.org/10.2307/2531691.

[13] Ibrahim, J. G., Chen, M. -H. and Sinha, D. (2001). Cure Rate Models. In *Bayesian Survival Analysis* 155–207. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4757-3447-8_5. MR1876598

[14] Ibrahim, J. G., Chen, M. and Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika* **88**(2) 551–564. https://doi.org/10.1093/biomet/88.2.551. MR1844851

[15] Kirkwood, J. M., Strawderman, M. H., Ernstoff, M. S., Smith, T. J., Borden, E. C. and Blum, R. H. (1996). Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the Eastern Cooperative Oncology Group Trial EST 1684. *Journal of Clinical Oncology* **14**(1) 7–17. https://doi.org/10.1200/JCO.1996.14.1.7.

[16] Lee, J. W. (1996). Some Versatile Tests Based on the Simultaneous Use of Weighted Log-Rank Statistics. *Biometrics* **52**(2) 721–725. https://doi.org/10.2307/2532911.

[17] Lee, S. H. (2007). On the versatility of the combination of the weighted log-rank statistics. *Computational Statistics and Data Analysis* **51**(12) 6557–6564. https://doi.org/10.1016/j.csda.2007.03.006. MR2408615

[18] Li, H., Tiwari, R. and Li, Q. H. (2022). Conditional borrowing external data to establish a hybrid control arm in randomized clinical trials. *Journal of Biopharmaceutical Statistics* **32**(6) 954–968. https://doi.org/10.1080/10543406.2021.2021227.

[19] Lin, R. S., Lin, J., Roychoudhury, S., Anderson, K. M., Hu, T., Huang, B., Leon, L. F., Liao, J. J. Z., Liu, R., Luo, X., Mukhopadhyay, P., Qin, R., Tatsuoka, K., Wang, X., Wang, Y., Zhu, J., Chen, T. -T. and Iacona, R. (2020). Alternative Analysis Methods for Time to Event Endpoints Under Nonproportional Hazards: A Comparative Analysis. *Statistics in Biopharmaceutical Research* **12**(2) 187–198. https://doi.org/10.1080/19466315.2019.1697738.

[20] Pepe, M. S. and Fleming, T. R. (1991). Weighted Kaplan-Meier Statistics: Large Sample and Optimality Considerations. *Journal of the Royal Statistical Society: Series B (Methodological)* **53**(2) 341–352. https://doi.org/10.1111/j.2517-6161.1991.tb01827.x. MR1108331

[21] Royston, P. and Parmar, M. K. (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Medical Research Methodology* **13**(1) 152. https://doi.org/10.1186/1471-2288-13-152.

[22] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B, Statistical methodology* **64**(4) 583–639. https://doi.org/10.1111/1467-9868.00353. MR1979380

[23] Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox Proportional Hazards Cure Model. *Biometrics* **56**(1) 227–236. https://doi.org/10.1111/j.0006-341X.2000.00227.x. MR1767631

[24] Yin, G. and Ibrahim, J. G. (2005). Cure Rate Models: A Unified Approach. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **33**(4) 559–570. https://doi.org/10.1002/cjs.5550330407. MR2232380

[25] Zhang, J., Guo, Y., Zhou, J. and Rasmussen, H. E. (2023). PubPredict: Prediction of progression and survival in oncology leveraging publications and early efficacy data. *Pharmaceutical Statistics* **22**(5) 963–973. https://doi.org/10.1002/pst.2321. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/pst.2321.

Hongfei Li. Department of Statistics, University of Connecticut, Storrs, CT, 06269, USA.
Incyte Corporation, Wilmington, DE, 19801, USA. E-mail address: hongfei.2.li@uconn.edu

Qian H. Li. StatsVita, LLC, Bethesda, MD, 20817, USA. E-mail address: qianhelenli@gmail.com

Ming-Hui Chen. Department of Statistics, University of Connecticut, Storrs, CT, 06269, USA. E-mail address: ming-hui.chen@uconn.edu