On Stability of Mixture Cure Models

Jia Bao* Xiaoxia Han† Yongzhao Shao‡

Abstract

With recent advancements in cancer research and other branches of medicine, many cancer patients can be clinically cured who will never experience disease recurrence or progression, or disease-specific death. In the presence of a high cure fraction, conventional survival models are not appropriate because they do not account for the possibility of cure. The mixture cure models (MCMs) have been developed with the EM based implementation to simultaneously estimate the cure fraction and the survival function of uncured patients. However, the available R packages for the EM-based implementation of the MCMs are lack of robustness, especially when the sample size is small. This paper investigates the stability of the estimates of the MCMs, and proposes a shrinkage EM algorithm for robust inference of mixture cure models by incorporating existing common knowledge on predictors as weakly informative priors. Numerical studies are conducted to show the instability of the ordinary EM-based estimates of MCMs and the advantages of shrinkage EM algorithm for robust inference of mixture cure models.

Key Words: Mixture cure models, EM algorithm, Bayesian prior, stability, shrinkage EM algorithm

1. Introduction

With the development of modern medicine and effective therapies, the curability of many cancers and other hard-to-treat diseases is becoming a reality. In early-stage melanoma and other diseases with good prognosis, frequently, many patients will never experience cancer recurrence in a long-term follow up. In the presence of cured subjects, the overall patient population consists of a mixture of cured and uncured subpopulations. In this context, the commonly used survival models, such as Cox proportional hazards model, are not ideal. These standard survial models assume all patients will eventually experience the event of interest given long enough follow up. In contrast, mixture cure models (MCMs) are particularly suitable for describing patient's survival and cure status. MCMs jointly model the latent cure status using logistic regression and model survival time using Cox regression models (Farewell 1982, Kuk and Chen 1992, Peng and Deer 2000, Sy and Taylor 2000, Han 2017, Zhang and Shao 2018). Importantly, Peng and Dear (2000) and Sy and Taylor (2000) applied the expectation–maximization (EM) algorithm to obtain the maximum likelihood estimators under the logistic-Cox mixture cure model. Despite extensive literature on MCMs and the associated conceptual flexibility, the MCMs has been under used. Cox regression models have been widely used even in the presence of both cured and uncured patients, partially due to easy-to-use software for the Cox regression models and unfamiliarity to the MCMs. Cai et al. (2012) introduced an R package *smcure* to to compute the maximum (partial) likelihood estimates of the semiparametric mixture cure models, which has became the most widely used R package for fitting MCMs.

As is well known, to identify the cure effect, it is important to have long follow up for some cured (eventually censored) subjects. In the case of limited study follow up and heavy censoring, there are stability issues for the EM based computation of the MCMs in

*Division of Biostatistics, NYU Grossman School of Medicine, New York, NY, 10016
†Henry Ford Hospital, Henry Ford Health System, Detroit, MI, 48202
‡Division of Biostatistics, NYU Grossman School of Medicine, New York, NY, 10016

smcure. To overcome this challenge, we suggest a shrinkage EM algorithm where we use the Bayesian GLM with default prior on the parameters of the logistic regression in the EM algorithm that can lead to more robust estimates.

In this paper, we develop a shrinkage EM algorithm to address issues associated with inferential stability due to latent heterogeneity in the form of unobserved cure status in the patient population. We also demonstrate the MCMs can easily include time-dependent predictors. Time-dependent predictors are important and frequently more useful than the baseline values. Numerical studies are used to demonstrate the advantages of the newly proposed shrinkage-based EM algorithms for robust inference.

2. Models

2.1 The Mixture Cure Models

Let T denote the failure time of interest and $S_{pop}(t \mid \boldsymbol{X}, \boldsymbol{Z})$ be the survival function of T depending on covariates $\boldsymbol{X}$ and $\boldsymbol{Z}$. The mixture cure model is defind as

$$S_{pop}(t \mid \boldsymbol{X}, \boldsymbol{Z}) = \pi(\boldsymbol{Z})S(t \mid \boldsymbol{X}) + 1 - \pi(\boldsymbol{Z}).$$

where $\pi(\boldsymbol{Z})$ is the probability of a patient being uncured depending on $\boldsymbol{Z}$, refered as "incidence" and $S(t|\boldsymbol{X})$ is the survival function of the failure time distribution of uncured patients depending on $\boldsymbol{X}$, referred as "latency". Usually, a logit link function is used to model the effects of covariates $\boldsymbol{Z}$. Define a binary variable Y as cure indicator, where $Y = 1$ indicates an individual will eventually experience the event and $Y = 0$ indicates an individual will never experience the event (or long-term survivor). Then the distribution of Y can be represented as a logistic model

$$Pr(Y = 1 \mid \boldsymbol{Z}) = \pi(\mathbf{Z}) = \frac{\exp(\mathbf{bZ})}{1 + \exp(\mathbf{bZ})},$$

where $\boldsymbol{b}$ is a vector of unknown parameters, is used to model the effect of $\boldsymbol{Z}$. Then $\pi(\boldsymbol{Z})$ is the probability of a patient being uncured, $1 - \pi(\mathbf{Z})$ is the cure probability.
When $S(t|\mathbf{X})$ is modeled by Cox PH model, that is $S(t \mid \boldsymbol{X}) = [S_0(t)]^{\exp(\beta^T \mathbf{X})}$, the above MCM model is called the logistic-Cox PH mixture cure model, where $S_0(t)$ is the unknown baseline survivial function. For brevity, we focus on this semiparametric logistic-Cox mixture cure models, which can be implemented by the *smcure* function in the *smcure* R package. However, *smcure* algorithm requires a relatively large sample size to give a reliable estimates. Otherwise, the EM algorithms may not converge or the variance of the estimates in logistic part may be inflated. Therefore, it is important to develop a robust procedure to improve stability in the EM-based estimation. In many applications, the sample size in clinical studies are not large, and even if the sample size is large but the follow up may be short as many studies or clinical trials have limitted study duration due to funding and other logistic constraints. Either sample size and/or short follow up can lead to instability of the implementation of the existing algorithms for MCMs such as Cai et al's R packge. Also, in Cai's *smcure* algorithm, latency part can only be modeled by fixed-time covariates. However, as is typical in many survival studies, individuals are monitored during the study, and some explanatory variables may change over time. Those time-dependent variables may be instrumental in predicting survival and need to be taken into consideration in evaluating the survival distribution.

2.2 Robust Implementation of MCMs

To address these problems of Cai's method, we develop robust mixture cure models (*rcure*) by incorporating Bayesian prior with availability to extend to time-dependent covariates. Our robust mixture cure model can be expressed as

$$S_{pop}(t \mid \mathbf{X}(t), \mathbf{Z}) = \pi(\mathbf{Z}) S(t \mid \mathbf{X}(t)) + 1 - \pi(\mathbf{Z}) \tag{1}$$

where $\pi(\boldsymbol{Z})$ is referred to as uncured probability depending on time-independent covariates $\boldsymbol{Z}$ which is modeled by a logistic model:

$$Pr(Y = 1|\mathbf{Z}) = \pi(\mathbf{Z}) = \frac{\exp(\mathbf{bZ})}{1 + \exp(\mathbf{bZ})}, \tag{2}$$

where unknown parameter $\boldsymbol{b}$ stands for the log odds ratio (OR). In real applications, we rarely have extremely informative predictors with OR larger than 10 or smaller than 0.1 per standard deviation for continuous predictors. Such predictors can be handled separately. Typical predictors have moderate ORs. Thus, as argued by Gelman et al (2008) and others, it is really without loss of generality to assume the ORs are moderate for the predictors in the above logistic regression. One way to implement this common knowledge is to assume the Log OR parameters in the above logistic model follow a weak prior distribution in the Bayesian GLM that can be implemented efficiently using existing R packages developed by Gelamn and colleagues. When the sample size is small, we suggest to use the default prior of a Cauchy distribution with zero location parameter and scale parameter 2.5.

$S(t \mid \mathbf{X}(t))$ denotes the survival probability of the uncured patients, where $\boldsymbol{X}(t)$ is the time-dependent covariates, whose values change over the duration of the follow-up time. The time-dependent Cox model is used to model the latency part, which can be expressed as:

$$\lambda(t \mid \boldsymbol{X}(t)) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{X}(t)), \tag{3}$$

where regression coefficient $\boldsymbol{\beta}$ is constant over time. $\lambda_0(t)$ is the baseline hazard function. Then, the survival function of uncure patients is given by

$$S(t \mid \mathbf{X}(t)) = \exp(-\int_0^t \lambda_0(u) \exp(\boldsymbol{\beta}^T \boldsymbol{X}(u)) du). \tag{4}$$

In order to obtain a robust estimation, we will introduce the shrinkage-based EM algorithm by incorporating a Bayesian GLM algorithm to implement the robust logistic regression as discussed in the next section.

3. Methods

3.1 Complete likelihood function of the MCMs

Let $\mathbf{O} = \{\boldsymbol{O}_1, \boldsymbol{O}_2, \ldots, \boldsymbol{O}_n\}$ denote the observed data. $\boldsymbol{O}_i = (t_i, \delta_i, \boldsymbol{z}_i, \boldsymbol{x}_i(t))$ denote the observed data for the ith individual, $i = 1, \ldots, n$, where t_i is the observed survival time, δ_i is the censoring indicator with $\delta_i = 1$ for uncensored time and $\delta_i = 0$ for censored time, $\mathbf{z}_i$ is the possible covariates in the incidence part, and $\mathbf{x}_i(t)$ is the time-dependent covariates in the latency part. The censoring is assumed noninformative and independent. Let $\boldsymbol{\Theta} = (\mathbf{b}, \boldsymbol{\beta}, \boldsymbol{\lambda}_0(t))$ denote the unknown parameters. Assume the latent cure indicator $\boldsymbol{Y} = \{y_1, y_2, \cdots, y_n\}$ is known, where $y_i = 1$ denotes the ith individual is not cured and $y_i = 0$ is cured, with probability $\pi(\boldsymbol{z}_i) = P(y_i = 1 \mid \boldsymbol{z}_i)$. The complete likelihood function is

$$\prod_{i=1}^{n}(1-\pi(\mathbf{z}_i))^{(1-y_i)}\pi(\mathbf{z}_i)^{y_i}\lambda(t_i \mid y_i=1,\mathbf{x}_i(t))^{\delta_i y_i}S(t_i \mid y_i=1,\mathbf{x}_i(t))^{y_i}. \tag{5}$$

Then the log likelihood function can be expressed as

$$l(\boldsymbol{b},\boldsymbol{\beta};\boldsymbol{O},\boldsymbol{Y}) = l_1(\boldsymbol{b};\boldsymbol{O},\boldsymbol{Y}) + l_2(\boldsymbol{\beta};\boldsymbol{O},\boldsymbol{Y}), \tag{6}$$

where

$$l_1(\boldsymbol{b};\boldsymbol{O},\boldsymbol{Y}) = \sum_{i=1}^{n}[y_i\log(\pi(\boldsymbol{z}_i)) + (1-y_i)\log(1-\pi(\boldsymbol{z}_i))], \tag{7}$$

$$l_2(\boldsymbol{\beta},\boldsymbol{\lambda}_0(t);\boldsymbol{O},\boldsymbol{Y}) = \sum_{i=1}^{n} y_i[\delta_i\log(\lambda(t_i \mid y_i=1,\boldsymbol{x}_i(t))) + \log(S(t_i \mid y_i=1,\boldsymbol{x}_i(t)))]. \tag{8}$$

Obviously, if $\delta_i = 1$, then $y_i = 0$, but if $\delta_i = 0$, y_i is not observable and y_i can be one or zero. Since Y is partially missing information, we do not have an exact expression for equation (8), the expectation maximization (EM) algorithm will be employed (Dempster, Laird, and Rubin 1977).

3.2 The Shrinkage EM algorithm

A shrinkage EM algorithm is proposed for our robust mixture cure models, to reduce the variance of the estimate in the logistic regression of mixture cure models. The E-step in the EM algorithm computes the conditional expectation of the log likelihood with respect to Y_i, given observed data $\boldsymbol{O}$ and current estimates of $\boldsymbol{\Theta}^{(m)}$. The conditional probability of the ith individual remaining uncured at the mth iteration of the algorithm can be written as

$$\begin{aligned} w_i^{(m)} =& E(y_i \mid \boldsymbol{O},\boldsymbol{\Theta}^{(m)}) \\ =& \delta_i + (1-\delta_i)\frac{\pi(\boldsymbol{z}_i)S(t_i \mid y_i=1,\boldsymbol{x}_i(t))}{1-\pi(\boldsymbol{z}_i)+\pi(\boldsymbol{z}_i)S(t_i \mid y_i=1,\boldsymbol{x}_i(t))}. \end{aligned} \tag{9}$$

It is easy to see that if $\delta_i = 1$, $w_i^{(m)} = 1$; and if $\delta_i = 0$, $w_i^{(m)}$ is the uncured probability of the ith individual. Then we have $\delta_i w_i^{(m)} = \delta_i$ and $\delta_i \log(w_i^{(m)}) = 0$. Since both parts of log likelihood function (7) and (8) are linear functions of y_i, the expectation of l_1 and l_2 can be written as

$$E(l_1) = \sum_{i=1}^{n}[w_i^{(m)}\log(\pi(\boldsymbol{z}_i)) + (1-w_i^{(m)})\log(1-\pi(\boldsymbol{z}_i))], \tag{10}$$

$$\begin{aligned} E(l_2) =& \sum_{i=1}^{n}[w_i^{(m)}\delta_i\log(\lambda(t_i \mid y_i=1,\boldsymbol{x}_i(t))) + w_i^{(m)}\log(S(t_i \mid y_i=1,\boldsymbol{x}_i(t)))] \\ =& \sum_{i=1}^{n}[\delta_i\log(\lambda(t_i \mid y_i=1,\boldsymbol{x}_i(t))) + w_i^{(m)}\log(S(t_i \mid y_i=1,\boldsymbol{x}_i(t)))] \\ =& \sum_{i=1}^{n}[\delta_i\log(w_i^{(m)}) + \delta_i\log(\lambda(t_i \mid y_i=1,\boldsymbol{x}_i(t))) + w_i^{(m)}\log(S(t_i \mid y_i=1,\boldsymbol{x}_i(t)))] \\ =& \sum_{i=1}^{n}[\delta_i\log(w_i^{(m)}\lambda(t_i \mid y_i=1,\boldsymbol{x}_i(t))) + w_i^{(m)}\log(S(t_i \mid y_i=1,\boldsymbol{x}_i(t)))]. \end{aligned} \tag{11}$$

The M-step in EM algorithm is to maximize the log likelihood function (6). The EM estimator can be obtained by maximizing equation (10) and (11). Gelman et al. (2008) developed a function *bayesglm* (part of the *arm* R package) by altering the *glm* function in R, which allows users to specify independent prior distribution for the coefficients in logistic regression models. Stable logistic regressions coefficient can be obtained by this Bayesian inference approach even when there is separation or sparsity in the dataset. Therefore, in order to get the robust estimate, we incoporate a default prior (Cauchy distribution with location parameter 0 and scale parameter 2.5) into the cure fraction of the mixture cure models. If informative prior is available, it is encouraged to be used instead of the default prior. Therefore, we use the function *bayesglm* in R to obtain robust estimate parameter b in equation (10). Peng and Dear (2000) and Sy and Taylor (2000) proposed a method to estimate $\boldsymbol{\beta}$ using partial likelihood. Then we can obtain estimate of $\boldsymbol{\beta}$ in equation (11) by estimating

$$\log \prod_{i=1}^{n} [\lambda_0(t_i) \exp(\boldsymbol{\beta} \boldsymbol{x}_i(t) + \log({w_i}^{(m)}))]^{\delta_i} S_0(t_i)^{\exp(\boldsymbol{\beta} \boldsymbol{x}_i(t) + \log({w_i}^{(m)}))}, \tag{12}$$

which is similar to the log-likelihood function of the commonly used time-dependent PH model with the additional offset variable $\log(w_i^{(m)})$. Thus, we can obtain the estimates in (12) by function *coxph* in R.

In order to proceed the E-step in the EM algorithm, we need to update the estimated survival function. Let $t_{(1)} < t_{(2)} < \cdots < t_{(k)}$ be the distict uncensored failure times, d_j denotes the number of events at time $t_{(j)}$ and R_j denotes the risk set include all individuals who are still under study at the time prior to $t_{(j)}$. The estimate for the baseline hazard function is

$$\lambda_{0j} = \hat{\lambda_0}(t \mid Y = 1) = \frac{d_j}{(t_{(j-1)} - t_{(j)}) \sum_{i \in R_j} w_i e^{\hat{\boldsymbol{\beta}} \boldsymbol{X}_i(t_{(j)})}} \quad \text{for } t \in (t_{(j-1)}, t_{(j)}]. \tag{13}$$

The baseline hazard function is assumed to be piecewise constant between failure times.

4. Simulation Studies

We use simulations to demenstrate the stability issue of the mixture cure models and incoporating Bayesian prior into logistic regression part can help to gain stability. Samples are generated from a logistic-Cox mixture cure model:

$$S_{pop}(t \mid X, Z) = \pi(Z) S(t \mid X) + 1 - \pi(Z).$$

Following Kuk and Chen (1992) and Peng and Dear (2000), we generate a control group of 30 observations and a treatment group of the same size, simple size n=60 in total. The indicator of the treatment group is the only covariate, denote as X. The probability of uncure is generated from a logistic model, where $\pi(Z) = \frac{\exp(b_0 + b_1 Z)}{1 + \exp(b_0 + b_1 Z)}$, with $Z = X$. The logistic parameters are set at $b_0 = 2$, $b_1 = -1$, so that the probability of uncured individuals $Pr(Y = 1 \mid Z = 0) = 0.881$ for the control group, and $Pr(Y = 1 \mid Z = 1) = 0.731$ for the treatment group, which means cure rate of 11.9% in the control group and 26.9% in the treatment group. The survival times are generated from a Cox model, where

$$S(t \mid Y = 1, X) = S_0(t \mid Y = 1)^{\exp(\beta X)}.$$

The parameter is set at $\beta = \log(1/2) = -0.693$. The standard exponential distribution is used for baseline survival funcution $S_0(t \mid Y = 1)$ for uncured patients in the control

group. The censoring times are generated from a exponetial distribution $\exp(\lambda_C)$, where $\lambda_C = 0.28$. This choice of λ_C gives an expected censoring proportion 31% for the control group, 53% for the treatment group, and overall is 42%. Under this setting, 500 samples are generated. Cai's *smcure* models and our proposed *rcure* models with default Cauchy (0, 2.5) prior are applied to the samples. As estimation results shown in Table 1, by incorporating the default prior, the estimated coefficients of the logistic regression part has much smaller bias, variance and MSE of our *rcure* algorithm compared to *smcure* algorithm. The bias for $\hat{b}_0$ by our method is around 10%, while the the bias for $\hat{b}_0$ by *smcure* is larger than 150%. The bias for $\hat{b}_1$ by our method is less than 5%, while the the bias for $\hat{b}_1$ by *smcure* is larger than 200%. In addition, our *rcure* methods give both the MSE of $\hat{b}_0$ and $\hat{b}_1$ less than 1, while the MSE of $\hat{b}_0$ and $\hat{b}_1$ by *smcure* are both around 50. On the other hand, in survival part, *smcure* algorithm and *rcure* algorithm both give stable and less biased $\hat{\beta}$. By our *rcure* algorithm, the bias of $\hat{\beta}$ is 3% and the MSE of $\hat{\beta}$ is 0.198; while by *smcure* algorithm, the bias of $\hat{\beta}$ is 2% and the MSE of $\hat{\beta}$ is 0.207.

Table 1: Comparison of *smcure* and *rcure* estimates
sample size n=60

		$\hat{b}_0$	$\hat{b}_1$	$\hat{\beta}$
	Model	2	-1	-0.693
Average	smcure	5.068	-3.269	-0.681
	rcure	2.224	-0.958	-0.715
Bias	smcure	3.068	-2.269	0.012
	rcure	0.224	0.042	-0.022
Variance	smcure	37.596	46.288	0.207
	rcure	0.802	0.953	0.198
MSE	smcure	47.009	51.436	0.207
	rcure	0.852	0.955	0.198

The instability of $\hat{b}_1$ by *smcure* can also be observed in the distribution plot of estimated b_1 in Figure 1. The blue bars represent the probability of estimated $\hat{b}_1$ obtained by *smcure* algorithm; the red bars represent the probability of estimated $\hat{b}_1$ obtained by *rcure* algorithm; the red vertical line represents the true value of b_1, $b_1 = -1$. As we can see from the upper plot, *smcure* gives quite wide and unstable estimates. The range of $\hat{b}_1$ by *smcure* algorithm can reach -20 and 20, while $\hat{b}_1$ by *rcure* algorithm is much more stable with range (-4.2, 1.9). The true value of b_1 is -1. However, by *smcure* algorithm, there are only 52.8% $\hat{b}_1$ fall in (-2,0), and 18.2% $\hat{b}_1$ even fall outside of (-10,10). We can even see a high frequency of $\hat{b}_1$ near -20 by *smcure*. There are 54 times $\hat{b}_1$ fall in (-20, -17) within 500 replications.

Then we conduct additional simulations for the same PH mixture cure model set up with larger sample size n=200 and same paramater setting as in Kuk and Chen (1992) and Peng and Dear (2000). Under this setting, 500 samples are generated. The *smcure* model and *rcure* model with default Cauchy (0, 2.5) prior are applied to the samples. Table 2 summarizes the average, biases and variances of the estimates of regression parameters from the *smcure* model and *rcure* model with defualt prior in the simulation study. When the sample size goes up to 200, the estimated b_0 and b_1 still have smaller bias and MSE by our *rcure* algorithm compared to Cai's *smcure* algorithm. The bias for $\hat{b}_0$ by our method is 4%, while the the bias for $\hat{b}_0$ by *smcure* is 20%. The bias for $\hat{b}_1$ by our method is less than 2%, while the the bias for $\hat{b}_1$ by *smcure* is 30%. In addition, the MSE of $\hat{b}_0$ and $\hat{b}_1$ by

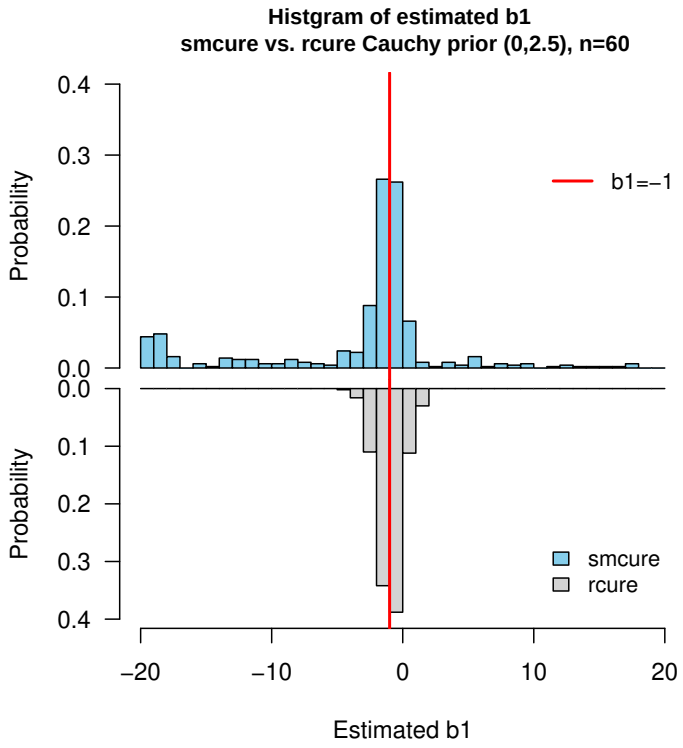

Figure 1: Histgram of $\hat{b}_1$, n=60

our method are both less than 0.5, while the MSE of $\hat{b}_0$ and $\hat{b}_1$ by *smcure* are both are over 3. For the survival part, stable and less biased $\hat{\beta}$ can be obtained by either Cai's *smcure* algorithm or our *rcure* algorithm. The two methods both give that the bias of $\hat{\beta}$ is less than 0.01 and the MSE of $\hat{\beta}$ is less than 0.1. Compared to the estimation results of sample size 60, both *smcure* algorithm and *rcure* algorithm get less bias and more stable estimates for all three parameters when the sample size goes up to 200.

When sample size is 200, the distribution of estimated b_1 by two mtheods can be compared in Figure 2. The blue bars represent the probability of estimated $\hat{b}_1$ obtained by *smcure* algorithm; the red bars represent the probability of estimated $\hat{b}_1$ obtained by *rcure* algorithm; the red vertical line represents the true value of b_1, $b_1 = -1$. The range of $\hat{b}_1$ by *smcure* is still wide, it is from -17.45 to 3.68. 83.4% $\hat{b}_1$ falls in (-2,0), while the true value of b is -1. However, our *rcure* algorithm gives shrinkage range (-3.3, 1,2), and nearly 90% $\hat{b}_1$ falls in (-2,0). Therefore, by incorporating the default prior, the mixture cure models can obtain a more stable and less biased slope estimate, especially when the sample size is small.

5. Discussion

The patient population often contains both cured and uncured patients in cancer research. Mixture cure models as an alternative to the conventional survival models, are useful to study survival of a patient population with a latent cure fraction. The *smcure* R package developed by Cai et al. (2012) is the main publicly available peer reviewed R package for fitting cure models. The *smcure* package adopts the conventional EM algorithm proposed by Peng and Dear (2000) and Taylor and Sy (2000) for mixture cure models. Compared to the conventional survival models, cure models have more parameters to be estimated.

Table 2: Comparison of *smcure* and *rcure* estimates
sample size n=200

		$\hat{b}_0$	$\hat{b}_1$	$\hat{\beta}$
	Model	2	-1	-0.693
Average	smcure	2.379	-1.300	-0.688
	rcure	2.082	-0.984	-0.703
Bias	smcure	0.379	-0.300	0.005
	rcure	0.082	0.016	-0.010
Variance	smcure	3.005	3.256	0.080
	rcure	0.270	0.4367	0.078
MSE	smcure	3.149	3.346	0.080
	rcure	0.277	0.436	0.078

Therefore, it requires a relatively large sample size and long follow-up to achieve stable estimation. In addition, in order to proceed with the E-step in EM algorithm, baseline hazard function need to be estimated. However, baseline hazard function generally has infinite dimensions, which is sometimes difficult to be estimated with a small sample size. A poorly estimated baseline hazard function may lead to unstable EM algorithm, and thus unstable parameter estimates in the logistic regression which describes cure status. Additionally, when sample size is small, the number of cured patients might be sparse causing the instability of logistic regression.

In this paper, we propose a shrinkage EM algorithm by incoporating Bayesian prior to obtain more stable estimate. The asymptotic validity of the Bayesian inference can be justified by the von Mise theorem (Le Cam 2012). As is well known, when the sample size is large, the maximum likelihood estimates are generally consistent under very general regularity conditions and asymptotically normal with minimal asymptotic variance as determined by the inverse of the Fisher information of the model. The von Mise theorem (Le Cam 2012) asserts that, under some general regularity conditions, the selection of prior distribution does not damage the efficiency of the posterior estimate which is asymptotically equivalent to the maximum likelihood estimates. Thus, formulating common-sense knowledge as prior in Bayesian GLM to implement the logistic regression will be justified if the sample size is large. Using good prior to gain stability and efficiency can also be justified from finite-sample consideration of the type of arguments similar to the von Mise theorem arguments. Assuming locally asymptotically normal (LAN) for the likelihood function, which is equivalent to say the likelihood is from an asymptotically normal model (Le Cam 2012), and assuming normal prior distribution, then the variance of the estimate using the posterior distribution is smaller than using either the observed data only or the prior only. Smaller variance in normal distribution means larger Fisher information, and easier estimation problems, and actually faster convergence of the MLEs and computing algorithms including the shrinkage EM algorithms. Indeed, the posterior likelihood is more spiky and more concave, and the original likelihood without using the common sense prior is more flat. Thus, the shrinkage EM algorithms leads to more robust estimates for mixture cure models as demonstrated by our simulation studies.

The same pattern also holds when we consider more general expamples including time-dependent covariates.

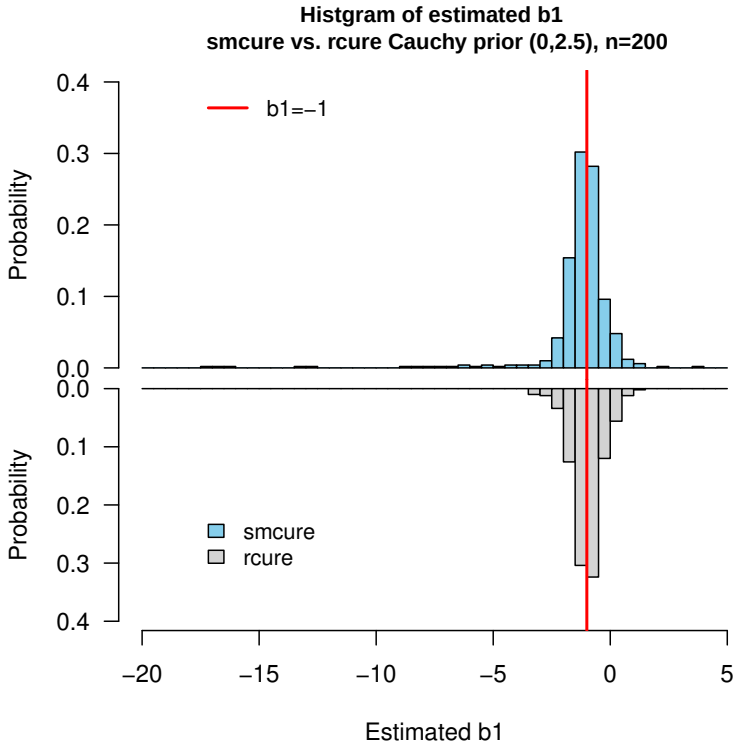

Figure 2: Histgram of $\hat{b}_1$, n=200

Acknowledgments

This research is partially supported by the NIH grants P50CA225450, P30CA016087, P30AG066512, P30ES000260, P01AG060882.

Conflict of Interest: None declared.

REFERENCES

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

Peng, Y., and Dear, K. B. (2000), "A nonparametric mixture model for cure rate estimation," *Biometrics*, 56(1), 237–243.

Sy, J. P., and Taylor, J. M. (2000), "Estimation in a Cox proportional hazards cure model," *Biometrics*, 56(1), 227–236.

A. Y. Kuk and C.-H. Chen (1992), "A mixture model combining logistic regression with proportional hazards regression," *Biometrika*, 531—541.

Han, X. (2017). "Statistical Methods for Analysis of Genetic and Survival Data with Latent Heterogeneity".

Zhang, Y. and Shao, Y. (2018). "Concordance measure and discriminatory accuracy in transformation cure models," *Biostatistics*, 19(1), 14–26.

Cai, C., Zou, Y., Peng, Y., and Zhang, J. (2012). "smcure: An R-Package for estimating semiparametric mixture cure models," *Computer methods and programs in biomedicine*, 108(3), 1255–1260.

C. Cai (2013), "Advanced methodology developments in mixture cure models".

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y. S. (2008), "A weakly informative default prior distribution for logistic and other regression models," *The annals of applied statistics*, 2(4), 1360–1383.

A. P. Dempster, N. M. Laird, and D. B. Rubin (1977), "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)*, 1-–38.

L. Le Cam (2012), *Asymptotic methods in statistical decision theory*, Springer Science & Business Media.