Published in final edited form as:
Comput Methods Programs Biomed. 2012 December ; 108(3): 1255–1260. doi:10.1016/j.cmpb.2012.08.013.

smcure: An R-package for Estimating Semiparametric Mixture Cure Models

Chao Cai[a], **Yubo Zou**[a], **Yingwei Peng**[b], and **Jiajia Zhang**[a,*]

[a]Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208, USA

[b]Department of Community Health and Epidemiology, Queen's University, Kingston, Ontario K7L 3N6, Canada

Abstract

The mixture cure model is a special type of survival models and it assumes that the studied population is a mixture of susceptible individuals who may experience the event of interest, and cure/non-susceptible individuals who will never experience the event. For such data, standard survival models are usually not appropriate because they do not account for the possibility of cure. This paper presents an R package smcure to fit the semiparametric proportional hazards mixture cure model and the accelerated failure time mixture cure model.

Keywords

Proportional hazards model; Accelerated failure time model; Semiparametric mixture cure model; EM algorithm; R package

1. Introduction

The proportional hazards (PH) model and the accelerated failure time (AFT) model are the most popular models in survival analysis. A common unstated assumption behind these models is that all patients will eventually experience the event of interest, given that the follow-up time is long enough. However, with the development of medical studies, more and more fatal diseases are now curable. Thus, there is a need to develop statistical models to analyze whether a treatment can cure the disease or slow down the progression of the disease if not curable. The mixture cure model, firstly introduced by Boag [2] and Berkson and Gage [1], is one of the most popular models to estimate the cure rate of treatment and the survival rate of uncured patients at the same time.

Let T denote the failure time of interest, $1 - \pi(\mathbf{z})$ be the probability of a patient being cured depending on $\mathbf{z}$, and $S(t|\mathbf{x})$ be the survival probability of uncured patients depending on $\mathbf{x}$,

*Corresponding author: jzhang@mailbox.sc.edu (Jiajia Zhang).

6. Availability

The package smcure and the relevant documentation can be freely downloaded from CRAN webpage http://cran.r-project.org/package=smcure.

where **x** and **z** are observed values of two covariate vectors that may affect the survival function. The mixture cure model can be expressed as

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

where $\pi(\mathbf{z})$ is referred to as "incidence" and $S(t|\mathbf{x})$ is referred to as "latency". If the PH model is used to model the latency part, the mixture cure model is called the proportional hazards mixture cure (PHMC) model. Instead, if the AFT model is applied to the latency distribution, it is called the accelerated failure time mixture cure (AFTMC) model.

In this paper, we present an R package named smcure to estimate the semiparametric PHMC and AFTMC models. In the next section, we outline the models and their computational methods. The R function and its arguments are described in Section 3. We use two examples to illustrate the smcure package in Section 4.

2. Models and Computational Methods

2.1. Semiparametric Mixture Cure Models

An advantage of the mixture cure models is that the proportion of cured subjects and the survival distribution of uncured subjects are modeled separately and the interpretation of effects of **x** and **z** is straightforward.

Usually, a logit link function

$$\pi(\mathbf{z})=\frac{\exp(\mathbf{bz})}{1+\exp(\mathbf{bz})},$$

where **b** is a vector of unknown parameters, is used to model the effects of **z**. Other link functions can also be applied to the incidence part, such as the complementary log-log link

$$\log(-\log(1-\pi(\mathbf{z})))=\mathbf{bz},$$

and the probit link

$$\Phi^{-1}(\pi(\mathbf{z}))=\mathbf{bz},$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. The logit link is a default option in the package.

As mentioned in the introduction, the latency part can be specified by the PH or the AFT model. Let $S_0(t)$ be the baseline survival function of uncured subjects when $\mathbf{x} = 0$. If $S(t|\mathbf{x}) = S_0(t)^{\exp(\boldsymbol{\beta}\mathbf{x})}$, the mixture cure model is called the PHMC model. If $S(t|\mathbf{x}) = S_0(te^{\boldsymbol{\beta}\mathbf{x}})$, it is the AFTMC model. Parametric approaches to the mixture cure models were studied by many authors [4, 9, 13]. Since it is usually difficult to verify a parametric assumption, there has been increasing interest in the semiparametric mixture cure models [7, 8, 11, 12, 14]. The smcure package in R will focus on semiparametric estimation methods for the PHMC and AFTMC models.

2.2. Computational Method

Let $\mathbf{O} = (t_i, \delta_i, \mathbf{z_i}, \mathbf{x_i})$ denote the observed data for the ith individual $i = 1, \cdots, n$, where t_i is the observed survival time, δ_i is the censoring indicator with $\delta_i = 1$ for the uncensored time and $\delta_i = 0$ for the censored time, and $\mathbf{z_i}$, $\mathbf{x_i}$ are the possible covariates in the incidence and latency parts respectively. We assume that the censoring is independent and noninformative. It is worthwhile to point out that the same covariates are allowed for the incidence and latency components although we use different covariate notations for these two components.

Let $\boldsymbol{\Theta} = (\mathbf{b}, \boldsymbol{\beta}, S_0(t))$ denote the unknown parameters. The EM algorithm is used to estimate the parameters of interest in the PHMC and AFTMC models. Let Y be the indicator that an individual will eventually ($Y = 1$) or never ($Y = 0$) experience the event, with the probability of $\pi(\mathbf{z})$. Given $\mathbf{y} = (y_1, y_2, \cdots, y_n)$ and $\mathbf{O}$, the complete likelihood function can be expressed as

$$\prod_{i=1}^{n} [1-\pi(\mathbf{z_i})]^{1-y_i} \pi(\mathbf{z_i})^{y_i} h(t_i|Y=1, \mathbf{x_i})^{\delta_i y_i} S(t_i|Y=1, \mathbf{x_i})^{y_i} \quad (2)$$

where $h(\cdot)$ is the hazard function corresponding to $S(\cdot)$. The logarithm of the complete likelihood function can be written as $l_c(\mathbf{b}, \boldsymbol{\beta}; \mathbf{O}, \mathbf{y}) = l_{c1}(\mathbf{b}; \mathbf{O}, \mathbf{y}) + l_{c2}(\boldsymbol{\beta}; \mathbf{O}, \mathbf{y})$, where

$$l_{c_1}(\mathbf{b}; \mathbf{O}, \mathbf{y}) = \sum_{i=1}^{n} y_i \log[\pi(\mathbf{z_i})] + (1-y_i)\log[1-\pi(\mathbf{z_i})], \quad (3)$$

$$l_{c_2}(\beta; \mathbf{O}, \mathbf{y}) = \sum_{i=1}^{n} y_i \delta_i \log[h(t_i|Y=1, \mathbf{x_i})] + y_i \log[S(t_i|Y=1, \mathbf{x_i})]. \quad (4)$$

The E-step in the EM algorithm computes the conditional expectation of the complete log-likelihood with respect to $y_i's$, given the observed data $\mathbf{O}$ and current estimates of parameters $\boldsymbol{\Theta}^{(\mathbf{m})} = (\mathbf{b}^{(m)}, \beta^{(m)}, S_0^{(m)}(t))$. The conditional expectation of y_i will be enough to complete this step since both (3) and (4) are linear functions of y_i. The expectation of $E(y_i|\mathbf{O}, \boldsymbol{\Theta}^{(\mathbf{m})})$ can be written as

$$w_i^{(m)} = E(y_i|\mathbf{O}, \boldsymbol{\Theta}^{(\mathbf{m})}) = \delta_i + (1-\delta_i) \frac{\pi(\mathbf{z}_i) S(t_i|Y=1, \mathbf{x_i})}{1-\pi(\mathbf{z_i}) + \pi(\mathbf{z}_i) S(t_i|Y=1, \mathbf{x_i})} \Big|_{(\mathbf{O}, \boldsymbol{\Theta}^{(\mathbf{m})})}.$$

It is easy to see that $w_i^{(m)} = 1$ if $\delta_i = 1$ and $w_i^{(m)}$ is the probability of uncured patients if $\delta_i = 0$. Thus, the second part of $E(y_i|\mathbf{O}, \boldsymbol{\Theta}^{(\mathbf{m})})$ can be interpreted as the conditional probability of the ith individual remaining uncured. Because $\delta_i \log w_i^{(m)} = 0$ and $\delta_i w_i^{(m)} = \delta_i$, the expectations of (3) and (4) can be written as

$$E(l_{c_1}) = \sum_{i=1}^{n} w_i^{(m)} \log[\pi(\mathbf{z_i})] + (1-w_i^{(m)}) \log[1-\pi(\mathbf{z_i})], \quad (5)$$

$$E(l_{c_2})=\sum_{i=1}^{n}\delta_i\log[w_i^{(m)}h(t_i|Y=1,\mathbf{x_i})]+w_i^{(m)}\log[S(t_i|Y=1,\mathbf{x_i})]. \quad (6)$$

The M-step in the EM algorithm is to maximize (5) and (6) with respect to the unknown parameters. We utilize the 'glm' function in R to estimate the parameters in equation (5). The 'link' option in the 'glm' can handle different link function in the mixture cure model.

Because the expressions of equation (6) and $w_i^{(m)}$ depend on the latency assumption, we will demonstrate the estimation approach under the PHMC model and the AFTMC model separately.

PHMC Model—Peng and Dear [10] and Sy and Taylor [11] proposed a partial likelihood type method to estimate $\boldsymbol{\beta}$ without specifying the baseline hazard function. The estimating equation (6) can be written as

$$\log\prod_{i=1}^{n}[h_0(t_i)\exp(\beta\mathbf{x}_i+\log(w_i^{(m)}))]^{\delta_i}S_0(t_i)^{\exp(\beta\mathbf{x}_i+\log(w_i^{(m)}))}, \quad (7)$$

which is similar to the log-likelihood function of the standard PH model with the additional offset variable $\log(w_i^{(m)})$. Therefore, we use the 'coxph' function in R to estimate the parameters in equation (6). A detailed presentation can be found in Peng [8], Peng and Dear [10], and Sy and Taylor [11].

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 distinct uncensored failure times, $d_{t_{(j)}}$ denote the number of events and $R(t_{(j)})$ denote the risk set at time $t_{(j)}$. The Breslow-type estimator for $S_0(t|Y=1)$ is given by

$$\widehat{S}_0(t|Y=1)=\exp\left(-\sum_{j:t_{(j)}\le t}\frac{d_{t_{(j)}}}{\sum_{i\in R(t_{(j)})}w_i^{(m)}e^{\widehat{\beta}\mathbf{x_i}}}\right). \quad (8)$$

Because the estimator, $\hat{S}_0(t|Y=1)$, may not approach 0 as $t\to\infty$, we set $\hat{S}_0(t|Y=1)=0$ for $t > t_{(k)}$. Then $\hat{S}(t|Y=1)=\hat{S}_0(t|Y=1)^{\exp(\hat{\boldsymbol{\beta}}\mathbf{x})}$.

AFTMC Model—Zhang and Peng [14] proposed a rank-based estimation method to estimate $\boldsymbol{\beta}$ in the M-step for the semiparametric AFTMC model. They turned equation (6) into a log-likelihood function of a standard semiparametric AFT model, except for the constant term $w_i^{(m)}$, which is

$$\log\prod_{i=1}^{n}[w_i^{(m)}h(\log(t_i)-\beta\mathbf{x}_i)]^{\delta_i}[S(\log(t_i)-\beta\mathbf{x}_i)^{w_i^{(m)}}].$$

This enables us to estimate $\boldsymbol{\beta}$ in the M-step by the existing semiparametric estimation methods for the AFT model [7]. Zhang and Peng [14] suggested to obtain the estimator by maximizing the convex function $G(\boldsymbol{\beta})$, where

$$G(\beta)=n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i w_j^{(m)}|\varepsilon_i-\varepsilon_j|I(\varepsilon_i-\varepsilon_j). \quad (9)$$

Therefore, maximization of (6) can be realized by maximizing (9) through the linear programming method in R.

Let $\tau_1 < \tau_2 < \cdots < \tau_k$ be the distinct uncensored failure residuals, which is log $t_i - \boldsymbol{\beta}\mathbf{x}_i$, $i = 1, \cdots, n$, d_{τ_j} denote the number of failures and $R(\tau_j)$ denote the risk set at τ_j. An estimator of $S_0(\varepsilon|Y=1)$ is given by

$$\widehat{S}_0(\varepsilon|Y=1)=\exp\left(-\sum_{j:\tau_j<\varepsilon}\frac{d_{\tau_j}}{\sum_{i\in R(\tau_j)}w_i^{(m)}}\right). \quad (10)$$

Same as the semiparametric PHMC model, we set $\hat{S}_0(\varepsilon|Y=1)=0$ for $\varepsilon > \tau_k$. Then $\hat{S}(t|Y=1)=\hat{S}_0(\varepsilon|Y=1)$.

2.3. Variance Estimation

Because of the complexity of the estimating equation in the EM algorithm, the standard errors of estimated parameters are not directly available. In order to obtain the variance of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{b}}$, this package randomly draws bootstrap samples with replacement by 'sample' function in R. The default number of bootstrap is set to be 100. We show that there is little difference in standard errors when using 100, 200 and 500 bootstrap samples in the illustrated examples (Tables 1 and 2), which means that 100 is enough for those two datasets.

3. Package Description

The estimation methods discussed above are implemented in the smcure package. The smcure function in the package can be called with the following syntax:

```
smcure(formula,cureform,offset=NULL,data,na.action=na.omit,model=c("ph",
"aft"),link="logit",Var=TRUE,emmax=50,eps=1e-7,nboot=100)
```

The required arguments are:

- formula: a formula object, with the response on the left of a '~' operator, and the variables included in the latency part on the right. The response must be a survival object as returned by the Surv function.
- cureform: specifies the variables included in the incidence part on the right of a '~' operator.
- data: a data frame containing variables used in formula and cureform.
- model: specifies survival model in the latency component, which can be "ph" or "aft".

The optional arguments are:

- offset: variable(s) with coefficient 1 in both incidence and latency parts of the semiparametric PHMC model or the semiparametric AFTMC model. By default, offset = NULL.
- na.action: a missing-data filter function. By default na.action = na.omit.
- link: specifies the link function in the incidence component. The logit, probit or complementary loglog (cloglog) links are available. By default link = “logit”.
- Var: if it is TRUE, the program returns bootstrap standard errors Std.Error for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{b}}$ by the bootstrap method. If it is set to be False, the program only returns coefficient estimates. By default, Var = TRUE.
- emmax: specifies the maximum iteration number. If the convergence criterion is not met, the EM iteration will be stopped after emmax iterations and the estimates will be based on the last maximum likelihood iteration. The default emmax = 50.
- eps: sets the convergence criterion. The default is eps = 1e-7. The iterations are considered to be converged when the maximum relative change in the parameters and likelihood estimates between iterations is less than the value specified.
- nboot: specifies the number of bootstrap samplings. The default nboot = 100.

The output is composed of two parts: Cure probability model and Failure time distribution model. The cure rate can be easily estimated from the output by $1-\hat{\pi}(\mathbf{z})$. The estimated mixture cure survival function $S_{pop}(\cdot)$ is computed by predictsmcure function and plotted by plotpredictsmcure function.

Remark

- For the categorical variable with more than two categories, say k categories, user has to create $k-1$ dummy variables outside the package.
- For the sake of identifiability, a covariate is required for both “formula” and “cureform” arguments.

4. Examples

In this section, we use two examples to illustrate the use of smcure package for the semiparametric PHMC model and the semiparametric AFTMC model respectively.

4.1. Eastern Cooperative Oncology Group (ECOG) Data

We fit the semiparametric PHMC model to the melanoma data from the ECOG phase III clinical trial e1684 [6], which was also illustrated by PSPMCM SAS macro [3]. The aim of the e1684 clinical trial was to evaluate the high dose interferon alpha-2b (IFN) regimen against the placebo as the postoperative adjuvant therapy. After deleting missing data, a total number of 284 observations is used in the analysis. Treatment (0=control,1=treatment), gender (0=male,1=female) and age (continuous variable which is centered to the mean) are used in both the incidence and latency parts. The response variable is relapse free survival in years. The semiparametric PHMC model can be fitted as following:

```
> pd <- smcure(Surv(FAILTIME,FAILCENS)~TRT+SEX+AGE,cureform=~TRT+SEX+AGE,
data=e1684,model=“ph”,nboot=500)
```

The output is:

```
> printsmcure(pd)
Call:
smcure(formula = Surv(FAILTIME, FAILCENS) ~ TRT + SEX + AGE, cureform =
~TRT + SEX + AGE, data = e1684, model = “ph”, nboot = 500, Var = TRUE)
Cure probability model:
Estimate Std.Error Z value Pr(>|Z|)
(Intercept) 1.36493298 0.28769252 4.7444159 2.091088e-06
TRT -0.58847727 0.30645148 -1.9202951 5.482064e-02
SEX -0.08696490 0.32905294 -0.2642885 7.915576e-01
AGE 0.02033857 0.01445227 1.4072922 1.593408e-01
Failure time distribution model:
Estimate Std.Error Z value Pr(>|Z|)
TRT -0.153595097 0.172120117 -0.8923716 0.3721938
SEX 0.099458470 0.190788176 0.5213031 0.6021556
AGE -0.007664013 0.006695195 -1.1447033 0.2523321
```

The standard errors of the estimated parameters are obtained based on 500 bootstrap samples. We also investigate the impact of the number of bootstrap samples on the standard error estimation. The estimated standard errors from different number of bootstrap samples (nboot) are listed in Table 1. We can see that there is little difference of standard error estimation.

If considering the male with the median centered age of 0.579, we can draw the fitted survival curves by the treatment group using the following commands:

```
> predm=predictsmcure(pd,newX=cbind(c(1,0),c(0,0),c(0.579,0.579)),
newZ=cbind(c(1,0),c(0,0),c(0.579,0.579)),model=“ph”)
> plotpredictsmcure(predm,model=“ph”)
```

The fitted survival curves for the male are shown in Figure 1(a). Similarly, we can fit the survival curves by the treatment group for the female at the same age, which are shown in Figure 1(b).

```
> predf=predictsmcure(pd,newX=cbind(c(1,0),c(1,1),c(0.579,0.579)),
newZ=cbind(c(1,0),c(1,1),c(0.579,0.579)),model=“ph”)
> plotpredictsmcure(predf,model=“ph”)
```

Both fitted survival curves show that the IFN treatment has higher survival probability than the placebo group.

4.2. Bone Marrow Transplant Study

To illustrate the semiparametric AFTMC model, we fit the bone marrow transplant study for the refractory acute lymphoblastic leukemia patients [5]. This data set is widely used in the AFTMC model because the PH assumption is not appropriate for the latency distribution [14]. There were 46 patients in the allogeneic treatment and 44 patients in the autologous treatment group. The treatment variable is included in both incidence and latency parts (1 for autologous treatment group; 0 for allogeneic treatment group). The semiparametric AFTMC model can be fitted as following:

```
> bmtfit <- smcure(Surv(Time,Status)~TRT,cureform=~TRT,
data=bmt,model="aft",nboot=200)
```

The output is:

```
> printsmcure(bmtfit)
Call:
smcure(formula = Surv(Time, Status) ~ TRT, cureform = ~TRT, data = bmt,
model = "aft", nboot = 200)
Cure probability model:
Estimate Std.Error Z value Pr(>|Z|)
(Intercept) 1.007354 0.2261408 4.4545448 8.407136e-06
TRT 0.427327 0.4843662 0.8822394 3.776474e-01
Failure time distribution model:
Estimate Std.Error Z value Pr(>|Z|)
(Intercept) 0.2101563 0.1783968 1.178027 0.2387859
TRT -0.3531250 0.2705977 -1.304982 0.1918991
```

The standard errors of estimated parameters are obtained based on 200 bootstrap samples. Again, we show the estimated standard errors under different number of bootstrap samples (nboot) in Table 2, and they are very close to each other.

The cure rate can be calculated based on the results from Cure probability model part. For example, the cure rate for the autologous transplant is 19.2 percent, which is calculated by $1 - \hat{\pi}(\mathbf{z}) = 1 - e^{1.007354+0.427327}/(1 + e^{1.007354+0.427327})$. The estimated survival curves with respect to the treatment can be obtained by

```
> predbmt=predictsmcure(bmtfit,newX=c(0,1),newZ=c(0,1),model="aft")
> plotpredictsmcure(predbmt,model="aft")
```

From the fitted survival curves in Figure 2, we can see that the patients from the allogeneic treatment group has better survival probability than those from the autologous treatment group.

5. Conclusions

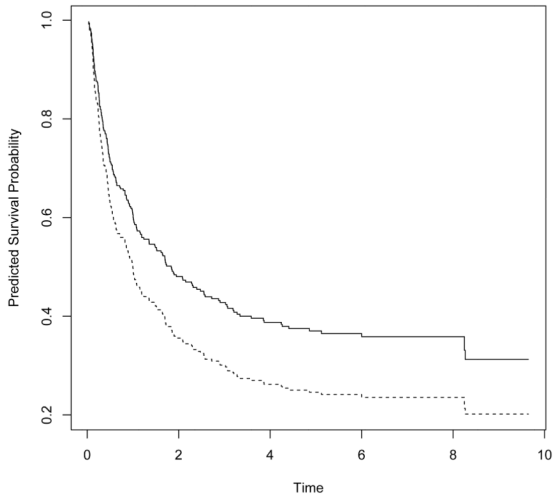
We develop an R package to estimate the semiparametric PHMC and AFTMC models. The cure probability part is estimated by the generalized linear model which allows many link functions, such as logit, probitand cloglog. The latency part can follow either the PH model or the AFT model. The semiparametric estimation procedures are based on the EM algorithm for both models. This package is an extension of the S-PLUS package **semicure** by Y. Peng which is for the PHMC model only, and the SAS macro PSPMCM [3] which accounts for the PHMC model and the parametric approach for the AFTMC model. The smcure package in R is developed for implementing the semiparametric estimation methods to both the PHMC model and the AFTMC model.

Acknowledgments

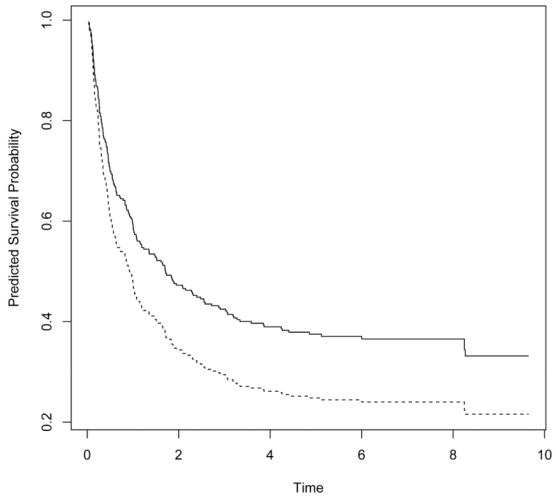
The project is supported by an award to Drs. Jiajia Zhang and Yingwei Peng from the National Cancer Institute (NCI, Award Number R03CA137790). The content is solely the responsibility of the authors and does not necessarily represent the official views of NCI.

References

1. Berkson J, Gage R. Survival curve for cancer patients following treatment. Journal of the American Statistical Association. 1952; 47:501–515.
2. Boag J. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. Journal of the Royal Statistical Society Series B (Methodological). 1949; 11(1):15–53.
3. Corbière F, Joly P. A sas macro for parametric and semiparametric mixture cure models. Computer methods and programs in biomedicine. 2007; 85(2):173–180. [PubMed: 17157948]
4. Farewell V. The use of mixture models for the analysis of survival data with long-term survivors. Biometrics. 1982; 38(4):1041–1046. [PubMed: 7168793]
5. Kersey J, Weisdorf D, Nesbit M, LeBien T, Woods W, McGlave P, Kim T, Vallera D, Goldman A, Bostrom B, et al. Comparison of autologous and allogeneic bone marrow transplantation for treatment of high-risk refractory acute lymphoblastic leukemia. New England Journal of Medicine. 1987; 317(8):461–467. [PubMed: 3302708]
6. Kirkwood J, Strawderman M, Ernstoff M, Smith T, Borden E, Blum R. Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the eastern cooperative oncology group trial est 1684. Journal of Clinical Oncology. 1996; 14(1):7. [PubMed: 8558223]
7. Li C, Taylor J. A semi-parametric accelerated failure time cure model. Statistics in medicine. 2002; 21(21):3235–3247. [PubMed: 12375301]
8. Peng Y. Fitting semiparametric cure models. Computational statistics & data analysis. 2003; 41(3): 481–490.
9. Peng Y, Dear K, Denham J, et al. A generalized f mixture model for cure rate estimation. Statistics in medicine. 1998; 17(8):813–830. [PubMed: 9595613]
10. Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. Biometrics. 2000; 56(1):237–243. [PubMed: 10783801]
11. Sy J, Taylor J. Estimation in a cox proportional hazards cure model. Biometrics. 2000; 56(1):227–236. [PubMed: 10783800]
12. Taylor J. Semi-parametric estimation in failure time mixture models. Biometrics. 1995; 51:899–907. [PubMed: 7548707]
13. Yamaguchi K. Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of “permanent employment” in japan. Journal of the American Statistical Association. 1992; 87:284–292.
14. Zhang J, Peng Y. A new estimation method for the semiparametric accelerated failure time mixture cure model. Statistics in medicine. 2007; 26(16):3157–3171. [PubMed: 17094075]



(a) Fitted survival curves for the male



(b) Fitted survival curves for the female

Figure 1.
Fitted survival curves for the male (a) and female (b) with median centered age by treatment groups for e1684 study. The upper solid line is the IFN treatment and lower dashed line is the control group.

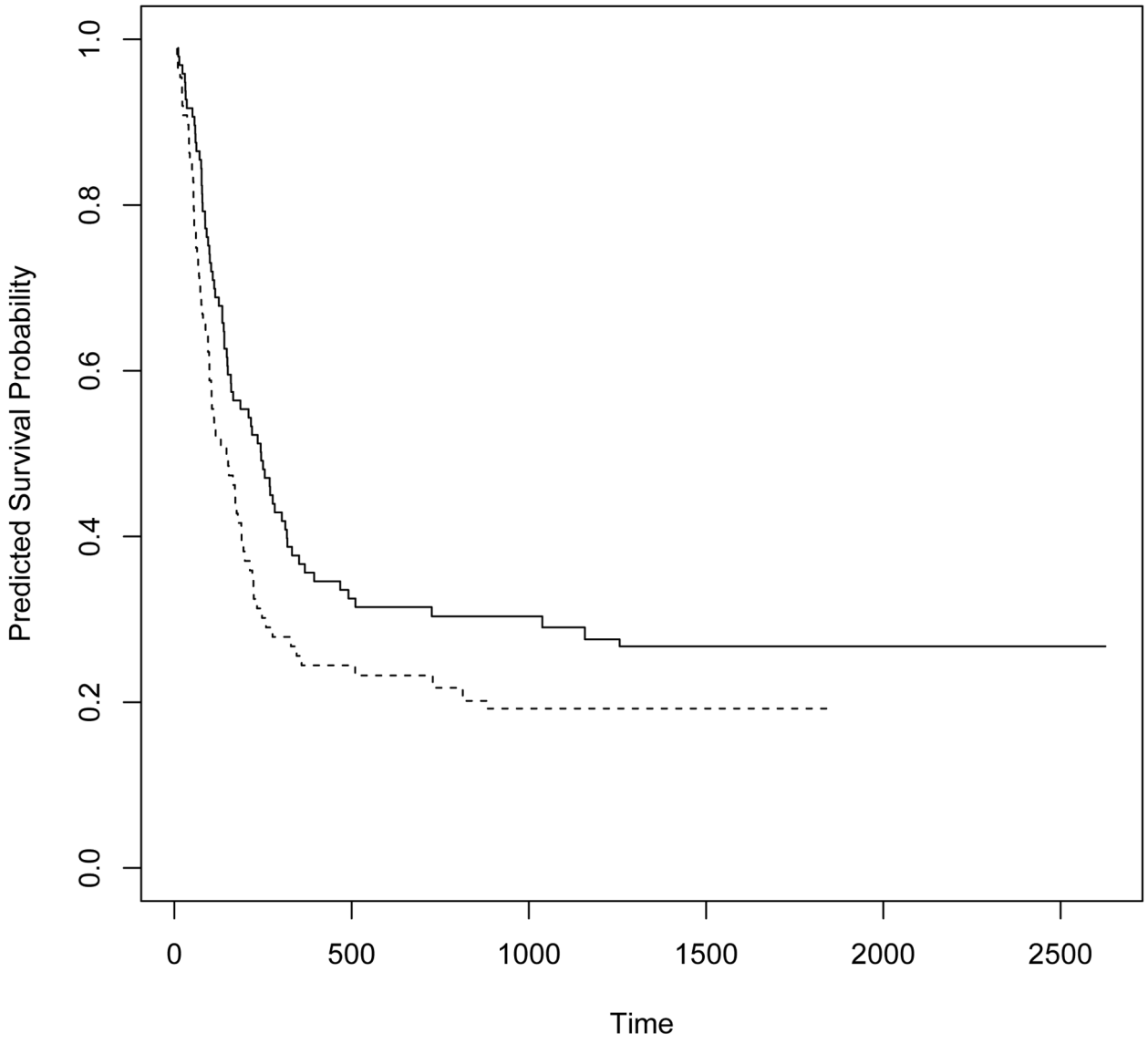


Figure 2.
Predicted Survival curves by treatment groups for bone marrow transplant study. The upper solid line is the allogeneic treatment group and lower dashed line is the autologous treatment group.

Table 1

Eastern Cooperative Oncology Group (ECOG) Data

Cure probability model	**SE(nboot=100)**	**SE(nboot=200)**	**SE(nboot=500)**
Intercept	0.35	0.33	0.29
TRT	0.36	0.33	0.31
SEX	0.34	0.33	0.33
AGE	0.02	0.01	0.01
Failure time distribution model	SE(nboot=100)	SE(nboot=200)	SE(nboot=500)
TRT	0.16	0.17	0.17
SEX	0.18	0.17	0.19
AGE	0.01	0.01	0.01

Table 2

Bone Marrow Transplant Study

Cure probability model	**SE(nboot=100)**	**SE(nboot=200)**	**SE(nboot=500)**
Intercept	0.25	0.23	0.26
TRT	0.52	0.48	0.54
Failure time distribution model	SE(nboot=100)	SE(nboot=200)	SE(nboot=500)
Intercept	0.21	0.18	0.18
TRT	0.30	0.27	0.27