Statistics and Its Interface Volume 7 (2014) 19–26

Semiparametric mixture survival model with application to MRFIT study

Zonghui Hu*, Jing Qin, and Dean A. Follmann

We study the mixture survival model where subject i has a probability p_i following one survival distribution and $1 - p_i$ following the other. The two survival distributions are unspecified except for an exponential tilting between the failure densities. Semiparametric likelihood estimation is proposed to handle censoring through conditional likelihood and inverse-censoring-probability weighted likelihood. Though full likelihood estimation is introduced, it is not always preferred over the other estimations due to its computational complexity and that its improvement in efficiency depends on the pattern of censoring. In the motivating example – MRFIT study, we apply mixture survival modeling to uncover the underlying survival patterns in the control arm: one for the would-be compliers and one for the would-be non-compliers, where compliance of each subject is not observable but associated with a probability.

Keywords and phrases: Censoring, Empirical likelihood, Exponential tilting, Inverse-censoring-probability weighting, Mixture distribution, Survival function.

1. INTRODUCTION

This paper investigates a mixture of two survival distributions with subject specific mixing probabilities. The motivating example comes from the Multiple Risk Factor Intervention Trial (MRFIT), see Multiple-Risk-Factor-Intervention-Trial-Group (1982; 1996). One objective of MRFIT was to study the effect of a special intervention, which involved counseling to quit smoking, on coronary heart disease mortality (Follmann, 2000). In that substudy, around 6,800 male participants who smoked were randomized to the special intervention (SI) arm and the usual care (UC) arm, and had a mean follow up of 7 years. An intention-to-treat (ITT) analysis, where patients were analyzed according to their original randomization arm regardless of their compliance, indicated no significant benefit from SI. Since only 23% participants under SI actually complied to the special intervention, it was suspected that the lack of benefit might be due to poor compliance. It is of interest whether compliance plays a role in the survival pattern of the participants. In the following, we define the "would-be" compliers, or "compliers" for simplicity, as the patients who would follow the intervention should they receive it. The goal of this work is to evaluate the difference in the survival pattern between the compliers and the non-compliers as well as the estimation of the survival patterns.

As in all clinical trials, compliers are identifiable in the treatment (SI) arm but not in the control (UC) arm. Nevertheless, each participant i in the UC arm has a tendency to comply should he receive SI; that is, he has a probability p_i to be a complier. This compliance probability is not directly observable. However, each person's compliance can be viewed as an inherent characteristics regardless of randomization to treatment or control (Efron & Feldman, 1991). Therefore, p_i is determined by the population compliance pattern and the participant's baseline characteristics. In Follmann (2000), the population compliance pattern was observed from the SI arm, and the baseline characteristics included the age, education, marital status, history and pattern of smoking, etc. Denote w as the vector of baseline characteristics, the compliance pattern $p(w)$ is fitted by a logistic regression of the compliance status versus w over the participants under SI. For subject i under UC, the probability of being a complier is $p_i = p(w_i)$ which we will treat as known thereafter.

There exist two underlying survival patterns in the UC arm: one for the complier and the other for the non-complier. Each subject follows a mixture survival pattern with failure density

$$f_i(t) = (1 - p_i)f_{.0}(t) + p_i f_{.1}(t), \tag{1}$$

with $f_{.0}$ standing for the failure density of the non-complier and $f_{.1}$ of the complier. Here, $p_i \in (0, 1)$ is known, $f_{.0}$ and $f_{.1}$ are unknown.

Our objective is to estimate survival function $S_{.0}(t) = \int_t^\infty f_{.0}(u)du$ and $S_{.1}(t) = \int_t^\infty f_{.1}(u)du$, and to evaluate the compliance effect on survival. Following the approach of Anderson (1979), we postulate exponential tilting between $f_{.1}(t)$ and $f_{.0}(t)$; that is,

$$f_{.1}(t) = \exp(\beta_0 + \beta_1 t)f_{.0}(t). \tag{2}$$

Model (1) together with (2) is a semiparametric mixture survival model. With exponential tilting (2), the compliance effect is reflected by β_1: $\beta_1 = 0$ corresponds to $S_{.1}(t) = S_{.0}(t)$ and $\beta_1 > 0$ corresponds to $S_{.1}(t) > S_{.0}(t)$. The advantage

*Corresponding author.

of the semiparametric model is its flexibility in data modeling. The exponential tilting (2) is true for $f_{.0}$ and $f_{.1}$ of exponential distributions or normal distributions with equal variances. If higher order terms of t are included, then (2) is true for $f_{.0}$ and $f_{.1}$ from exponential distribution families and approximately true for other distribution families. Without the exponential tilting, (1) is totally nonparametric and unidentifiable.

Mixture distribution models of form (1) have been studied by other researchers. Some work focuses on estimation of the mixing probabilities p_i, see Hall & Titterington (1984), McLachlan & Basford (1988), Qin (1999), and Qin & Leung (2005). Other works effectively assume p_i's known and estimate the component distributions, see Follmann (2000) and Ma et al. (2011). In this work, we apply mixture modeling to survival data. A complication with survival data is the incomplete observation from censoring. We propose semiparametric likelihood estimations via empirical likelihood (Owen, 1988; Qin & Lawless, 1994) and take care of censoring through conditional likelihood and inverse-censoring-probability weighted likelihood. We also introduce the full likelihood estimation through the Expectation-Maximization algorithm (Dempster et al., 1977), which is not always preferred due to its computational complexity and its limited improvement in efficiency under some censoring patterns.

The outline of the paper follows. In section 2, we propose semiparametric likelihood estimation for model (1) with (2). We explore their asymptotic properties in section 3 and investigate the numerical properties in section 4. In section 5, we apply the proposed model to the MRFIT study. In section 6 are the concluding remarks.

2. SEMIPARAMETRIC LIKELIHOOD ESTIMATION

We use $\{(T_i, C_i, \Delta_i), i = 1, \ldots, N\}$ for the full data, where T_i is the time of event, C_i the time of censoring, and $\Delta_i = I(T_i \leq C_i)$ the indicator of non-censoring. The observed data is $\{(X_i, \Delta_i), i = 1, \ldots, N\}$ with $X_i = \min(T_i, C_i)$. Let $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ and $\mathbf{t} = (1, t)'$, we express the exponential tilting (2) as $\exp(\boldsymbol{\beta}'\mathbf{t})$. The estimation and the properties stay the same for exponential tilting with higher order terms of t.

2.1 Conditional empirical likelihood

Let $f^*_{.0}(t) = f_{.0}(t \mid \delta = 1)$ and $f^*_{.1}(t) = f_{.1}(t \mid \delta = 1)$ stand for the conditional failure densities over the uncensored. Thus

$$(3)\quad f^*_{.0}(t) = \frac{f_{.0}(t)G(t)}{\int f_{.0}(t)G(t)dt} \quad \text{and} \quad f^*_{.1}(t) = \frac{f_{.1}(t)G(t)}{\int f_{.1}(t)G(t)dt},$$

where $G(t) = \mathrm{pr}(\delta = 1 \mid t)$ is the survival function of censoring. Following (1) and (3) and with $\theta = \int f_{.0}(t)G(t)dt / \int f_{.1}(t)G(t)dt$, we have

$$f_i(t \mid \delta = 1) = \frac{(1-p_i)\theta f^*_{.0}(t) + p_i f^*_{.1}(t)}{(1-p_i)\theta + p_i}.$$

The conditional density $f_i(t \mid \delta = 1)$ is thus a mixing of $f^*_{.0}(t)$ and $f^*_{.1}(t)$. Here we make the implicit assumption that all subjects follow the same censoring distribution $G(t)$. It follows from (2) and (3) that

$$f^*_{.1}(t) = \theta \exp(\boldsymbol{\beta}'\mathbf{t}) f^*_{.0}(t).$$

The conditional likelihood is $L_c = \prod_{\delta_i=1} f_i(t_i \mid \delta_i = 1)$. We maximize L_c using empirical likelihood (Owen, 1988; Qin & Lawless, 1994). Without loss of generality, suppose the first n subjects are the uncensored. Let $\mathbf{q}^*_0 = (q^*_{01}, \ldots, q^*_{0n})$ with $q^*_{0i} = f^*_{.0}(t_i)$ and t_i the event time observed from the i-th uncensored subject, then $f^*_{.1}(t_i) = \theta \exp(\boldsymbol{\beta}'\mathbf{t}_i) q^*_{0i}$. The estimate of $(\theta, \boldsymbol{\beta}, \mathbf{q}^*_0)$ is the maximizer of the log-likelihood

$$l_c(\theta, \boldsymbol{\beta}, \mathbf{q}^*_0) = \sum_{i=1}^{n} \left(\log\left\{ \frac{(1-p_i) + p_i \exp(\boldsymbol{\beta}'\mathbf{t}_i)}{(1-p_i)\theta + p_i} \right\} + \log(q_{0i}) \right) + n\log(\theta),$$

subject to the constraints

$$\sum_{i=1}^{n} q^*_{0i} = 1, \quad \sum_{i=1}^{n} \theta \exp(\boldsymbol{\beta}'\mathbf{t}_i) q^*_{0i} = 1.$$

These constraints ensure that $f^*_{.0}(t)$ and $f^*_{.1}(t)$ are density functions; that is $\int f^*_{.0}(t)dt = 1$ and $\int f^*_{.1}(t)dt = 1$. With $(\theta, \boldsymbol{\beta})$ fixed, maximization of l_c over $\mathbf{q}^*_0$ under the constraints gives

$$q^*_{0i} = n^{-1} \frac{1}{1 + \rho\{\theta \exp(\boldsymbol{\beta}'\mathbf{t}_i) - 1\}}$$

where ρ is the Lagrange multiplier

$$\rho = \rho(\theta) = n^{-1} \sum_{i=1}^{n} \frac{p_i}{(1-p_i)\theta + p_i}.$$

The profile log-likelihood of $(\theta, \boldsymbol{\beta})$ is then

$$l_{pc}(\theta, \boldsymbol{\beta}) = \sum_{i=1}^{n} \left(\log \frac{(1-p_i) + p_i \exp(\boldsymbol{\beta}'\mathbf{t}_i)}{(1-p_i)\theta + p_i} - \log[1 + \rho\{\theta \exp(\boldsymbol{\beta}'\mathbf{t}_i) - 1\}] \right) + n\log(\theta).$$

The estimate $(\theta, \boldsymbol{\beta})$ is the maximizer of l_{pc}. If we denote the estimates as $\widehat{\theta}, \widehat{\boldsymbol{\beta}}$ and let $\widehat{\rho} = \rho(\widehat{\theta})$, the estimates of the conditional failure densities are

$$\widehat{q}^*_{0i} = \widehat{f^*_{.0}}(t_i) = n^{-1}\frac{1}{1+\widehat{\rho}\{\widehat{\theta}\exp(\widehat{\boldsymbol{\beta}}'\mathbf{t}_i)-1\}},$$

$$\widehat{q}^*_{1i} = \widehat{f^*_{.1}}(t_i) = n^{-1}\frac{\exp(\widehat{\boldsymbol{\beta}}'\mathbf{t}_i)}{1+\widehat{\rho}\{\widehat{\theta}\exp(\widehat{\boldsymbol{\beta}}'\mathbf{t}_i)-1\}}.$$

From (3), the unconditional failure densities can be estimated by

$$\widehat{f_{.0}}(t_i) = \gamma_0\widehat{f^*_{.0}}(t_i)/\widehat{G}(t_i), \quad \widehat{f_{.1}}(t_i) = \gamma_1\widehat{f^*_{.1}}(t_i)/\widehat{G}(t_i)$$

where $\widehat{G}(t)$ is the Kaplan-Meier estimate of $G(t)$ treating event as censoring and censoring as event, $\gamma_0 = \{\sum_{i=1}^n \widehat{q}^*_{0i}/\widehat{G}(t_i)\}^{-1}$ and $\gamma_1 = \{\sum_{i=1}^n \widehat{q}^*_{1i}/\widehat{G}(t_i)\}^{-1}$. The semiparametric estimates of the survival functions are

$$\widehat{S}_{.0}(t) = \left\{\sum_{i=1}^n \widehat{q}^*_{0i}/\widehat{G}(t_i)\right\}^{-1}\sum_{i=1}^n \frac{\widehat{q}^*_{0i}}{\widehat{G}(t_i)}I(t_i>t),$$

$$\widehat{S}_{.1}(t) = \left\{\sum_{i=1}^n \widehat{q}^*_{1i}/\widehat{G}(t_i)\right\}^{-1}\sum_{i=1}^n \frac{\widehat{q}^*_{1i}}{\widehat{G}(t_i)}I(t_i>t).$$

2.2 Inverse-censoring-probability weighted empirical likelihood

For an observation (x_i,δ_i) with $\delta_i = 0$, the event time t_i is missing due to censoring. Therefore, the probability of observing the event time is $G(t_i)$, the survival function of censoring. We can adopt the approach of inverse probability weighting for missing data (Horvitz & Thompson, 1952) to the estimation of mixture survival. The inverse-censoring-probability weighted likelihood is

$$l_w = \sum_{i=1}^N \frac{\delta_i \log\{f_i(t_i)\}}{G(t_i)}.$$

If there is no censoring, $\delta_i = G(t_i) = 1$ for all subjects and l_w is the full likelihood function. Again, we apply empirical likelihood estimation. Let $\mathbf{q}_0 = (q_{01},\dots,q_{0N})$ with $q_{0i} = f_{.0}(t_i)$, then $f_{.1}(t_i) = \exp(\boldsymbol{\beta}'\mathbf{t}_i)q_{0i}$. Denote $g_i = G(t_i)$, the weighted log-likelihood has the form

$$l_w(\boldsymbol{\beta},\mathbf{q}_0) = \sum_{i=1}^N \delta_i/g_i\left[\log\{(1-p_i)+p_i\exp(\boldsymbol{\beta}'\mathbf{t}_i)\}+\log(q_{0i})\right].$$

The estimate of $(\boldsymbol{\beta},\mathbf{q}_0)$ is the maximizer of l_w subject to

$$\sum_{i=1}^N q_{0i} = 1, \quad \sum_{i=1}^N q_{0i}\exp(\boldsymbol{\beta}'\mathbf{t}_i) = 1,$$

where the first constraint corresponds to $\int f_{.0}(t)dt = 1$ and the second to $\int f_{.1}(t)dt = 1$. With $\boldsymbol{\beta}$ fixed, maximization of l_w over $\mathbf{q}_0$ under the constraints gives

$$q_{0i} = N_w^{-1}\frac{\delta_i}{g_i}\frac{1}{1+\rho\{\exp(\boldsymbol{\beta}'\mathbf{t}_i)-1\}},$$

where $N_w = \sum_{i=1}^N \delta_i/g_i$, and ρ is the Lagrange multiplier

$$\rho = \rho(\boldsymbol{\beta}) = N_w^{-1}\sum_{i=1}^N \frac{\delta_i}{g_i}\frac{p_i\exp(\boldsymbol{\beta}'\mathbf{t}_i)}{(1-p_i)+p_i\exp(\boldsymbol{\beta}'\mathbf{t}_i)}.$$

The profile log-likelihood of $\boldsymbol{\beta}$ is then

$$l_{pw}(\boldsymbol{\beta}) = \sum_{i=1}^N \frac{\delta_i}{g_i}\Big(\log\{(1-p_i)+p_i\exp(\boldsymbol{\beta}'\mathbf{t}_i)\} - \log[1+\rho\{\exp(\boldsymbol{\beta}'\mathbf{t}_i)-1\}]\Big).$$

We estimate $\boldsymbol{\beta}$ as the maximizer of l_{pw}. Denote the estimate as $\widehat{\boldsymbol{\beta}}$ and let $\widehat{\rho} = \rho(\widehat{\boldsymbol{\beta}})$. The estimates of the failure densities are

$$\widehat{q}_{0i} = \widehat{f_{.0}}(t_i) = N_w^{-1}\frac{\delta_i}{g_i}[1+\widehat{\rho}\{\exp(\widehat{\boldsymbol{\beta}}'\mathbf{t}_i)-1\}]^{-1},$$

$$\widehat{q}_{1i} = \widehat{f_{.1}}(t_i) = N_w^{-1}\frac{\delta_i}{g_i}[1+\widehat{\rho}\{\exp(\widehat{\boldsymbol{\beta}}'\mathbf{t}_i)-1\}]^{-1}\exp(\widehat{\boldsymbol{\beta}}'\mathbf{t}_i)$$

for $i = 1,\dots,N$. It is obvious that $\widehat{q}_{0i} = \widehat{q}_{1i} = 0$ if $\delta_i = 0$. The semiparametric estimates of the survival functions are $\widehat{S}_{.0}(t) = \sum_{i=1}^N \widehat{q}_{0i}I(t_i>t)$ and $\widehat{S}_{.2}(t) = \sum_{i=1}^N \widehat{q}_{1i}I(t_i>t)$.

In the implementation of the weighted likelihood estimation, $g_i = G(t_i)$ is replaced by its Kaplan-Meier estimate $\widehat{g}_i$. According to the properties of inverse probability weighting for missing data, using $\widehat{g}_i$ instead of g_i does not affect consistency of the weighted estimator (Horvitz & Thompson, 1952; Wang et al., 1998).

2.3 Full empirical likelihood

We can construct a full likelihood involving both the censored and the uncensored observations. We use $O = \{(x_i,\delta_i), i = 1,\dots,N\}$ for the observed data and $t_1,\dots,t_n$ for the observed event times. Let T_i stand for the event time of subject i, then $T_i = x_i$ if $\delta_i = 1$ and $T_i > x_i$ is not observed if $\delta_i = 0$. The full log-likelihood can be written as

$$l_f = \sum_{j=1}^n\sum_{i=1}^N I(T_i = t_j)\log\{f_i(t_j)\},$$

which is not evaluable due to censoring. We adopt the Expectation-Maximization (EM) algorithm to evaluate $l_e = E(l_f \mid O)$. We see that

$$E\{I(T_i = t_j) \mid O\} = \delta_i I(x_i = t_j) + (1-\delta_i)\mathrm{pr}(T_i = t_j \mid T_i \geq x_i, O),$$

where the conditional probability is $\mathrm{pr}(T_i = t_j, x_i \leq t_j \mid O)/\mathrm{pr}(T_i \geq x_i \mid O)$ and equals $f_i(t_j)I(x_i \leq t_j)/\int_{x_i}^\infty f_i(t)dt$. The expected log-likelihood thus takes the form

$$l_e = \sum_{j=1}^n\sum_{i=1}^N w_{ij}\log\{f_i(t_j)\},$$

with

$$w_{ij} = \delta_i I(x_i = t_j) + (1-\delta_i)\frac{f_i(t_j)I(x_i \le t_j)}{\sum_{l=1}^n f_i(t_l)I(x_i \le t_l)}.$$

For a subject with $\delta_i = 1$, $w_{ij} = 1$ and $f_i(t)$ has a point mass at t_j for some $j \in \{1,\ldots,n\}$. For a subject with $\delta_i = 0$, $w_{ij} > 0$ if and only if $t_j \ge x_i$; that is, $f_i(t)$ has point masses over those t_j's to the right of x_i. This coincides with the "redistribution to the right" algorithm in Kaplan-Meier estimation (Efron, 1967). Instead of uniform redistribution adopted in Kaplan-Meier, the expected full likelihood l_e redistributes the weights differentially over $\{t_l : t_l \ge x_i, l = 1,\ldots,n\}$ as $f_i(t)$ changes with t.

Letting $\mathbf{q}_0 = (q_{01},\ldots,q_{0n})$ with $q_{0j} = f_{.0}(t_j)$, l_e can be written as

(4)

$$l_e(\boldsymbol{\beta},\mathbf{q}_0) = \sum_{j=1}^{n}\sum_{i=1}^{N} w_{ij}\left[\log\{(1-p_i)+p_i\exp(\boldsymbol{\beta}'\mathbf{t}_j)\}+\log(q_{0j})\right].$$

The estimate of $(\boldsymbol{\beta},\mathbf{q}_0)$ is the maximizer of (4) subject to constraints $\sum_{j=1}^n q_{0j} = 1$ and $\sum_{j=1}^n q_{0j}\exp(\boldsymbol{\beta}'\mathbf{t}_j) = 1$. We can use a two-step iterative EM algorithm for the estimation, see the Appendix.

3. PROPERTIES OF SEMIPARAMETRIC ESTIMATORS

The semiparametric likelihood estimates of $\widehat{\boldsymbol{\beta}}$, $\widehat{S}_{.0}(t)$, and $\widehat{S}_{.1}(t)$ are consistent under regularity conditions, see the Appendix. Due to profiling in the semiparametric likelihood estimations, the asymptotic variances are hard to derive and we recommend bootstrapping for variance estimation.

Both the conditional likelihood and the weighted likelihood estimators are constructed over the uncensored observations. Since l_c in section 2.1 is a likelihood function, the conditional likelihood estimator is expected to be at least as efficient as the weighted estimator. The weighted estimator can be quite variable when there are very low estimates of $G(t)$ at some uncensored observations, which is a common reservation for inverse weighting estimation. Except for that, the weighted estimator has better finite sample efficiency than the conditional likelihood estimator for two reasons. First, the weighted estimator additionally utilizes the censoring information as estimation of $G(t)$ includes both the uncensored and the censored observations. Second, the conditional likelihood has one more parameter θ than the weighted likelihood. We will see that, as sample size gets larger, impact from the two aspects becomes negligible and the conditional likelihood estimator shows better efficiency.

The full likelihood estimator may improve efficiency over the conditional likelihood estimator and the weighted likelihood estimator, with the gain of efficiency through redistributing the point mass at a censoring time over the event times to the right. Improvement in efficiency depends on the amount as well as the pattern of censoring. Suppose subject i is censored. If there is no event observed after x_i, then $w_{ij} = 0$ for all $j = 1,\ldots,n$ in the likelihood function l_e, and this censored subject does not contribute to the estimation. Therefore, when censoring occurs late in followup, gain of efficiency is limited. This scenario can happen in clinical trials where most subjects are censored due to the close of study. Furthermore, the iterative EM algorithm is computationally intensive with slower convergence than the other two, and it can be sensitive to initial value selection. Thus, full likelihood estimation is not always the preferred one.

In the following, we evaluate the numerical performance of the three estimators by simulations.

4. SIMULATION STUDIES

Let the survival density $f_i(t)$ be a mixture of two components as in (1), where we take $f_{.0}(t) = \exp(-t/\lambda)/\lambda$ with $\lambda = 10$ and $f_{.1}(t) = f_{.0}(t)\exp(\boldsymbol{\beta}'\mathbf{t})$ with $\beta_1 = -0.1$. It follows that $\beta_0 = 0.69$ and $f_{.1}(t)$ is exponential with mean $1/(\lambda^{-1} - \beta_1)$. Finally, let the censoring time follow exponential distribution with mean λ_c. Let the mixture probability, p_i for $i = 1,\ldots,N$, follow identical and independent $Beta(1,1)$.

We explore the numerical performances of the semiparametric likelihood estimations. In addition, we compute the parametric likelihood estimates, where $f_{.0}$ is correctly specified as exponential but with $\lambda, \beta_0, \beta_1$ estimated. For all estimations, we report the estimates of β_1 and the survival functions $S_{.0}(t)$.

Table 1 presents the estimates at $\lambda_c = 30$ which corresponds to around 20% censoring. We see root-n consistency in the estimates of β_1 and $S_{.0}(t)$. With 20% censoring, the full likelihood estimator shows only slight improvement in efficiency over the other estimators. At small and moderate sample sizes of 400 and 1,600, the weighted likelihood estimator has better performance than the conditional likelihood estimator. At sample size of 6,400, all three estimators are quite close.

Table 2 presents the estimates at $\lambda_c = 8$ which corresponds to around 50% censoring. Again, we observe root-n consistency in all the estimates. At large sample sizes of 3,200 and 12,800, the conditional likelihood estimation shows a better efficiency than the weighted likelihood estimator. At a high censoring rate of 50%, the full likelihood estimator shows more gain in efficiency than in Table 1.

As discussed in section 3, the pattern of censoring affects the relative efficiency of the full likelihood estimator over the others. In Table 3, we present the estimates of β_1 under the same setup for Table 2 except that all subjects have the same censoring time at $C_i = 5$. Though it corresponds roughly to a 50% censoring, the full likelihood estimation shows no improvement in efficiency. In this extreme scenario, the censored has no contribution in the full likelihood estimation.

Table 1. Estimates at sample size $N = 400$, 1,600, and 6,400 with 20% censoring: t_1, t_2, and t_3 are the 10%, 50%, and 90%-th percentiles of $f_{\cdot 0}(t)$. Bias is 100× the Monte-Carlo bias, SD* is 100× the Monte-Carlo standard deviation, and RE is the relative efficiency as the ratio of the SD of the full likelihood estimator versus the SD of the other two estimators*

N		Conditional			Weighted			Full	
		Bias*	SD*	RE	Bias*	SD*	RE	Bias*	SD*
400	β_1	−2.9	5.2	0.8	−0.7	4.9	0.8	−0.6	4.1
	$S_{\cdot 0}(t_1)$	0.8	3.1		−0.1	2.2		0	2
	$S_{\cdot 0}(t_2)$	2.6	8.7		−0.7	5.8		−0.1	4.8
	$S_{\cdot 0}(t_3)$	1.3	4.6		−1.4	2.9		−0.5	2.9
1,600	β_1	−0.2	2.5	0.8	−0.1	2.2	0.9	−0.1	1.9
	$S_{\cdot 0}(t_1)$	−0.3	1.5		0	1		0	0.9
	$S_{\cdot 0}(t_2)$	−0.9	3.8		−0.1	2.6		0	2.4
	$S_{\cdot 0}(t_3)$	−0.5	2.3		−0.4	1.5		−0.1	1.0
6,400	β_1	−0.1	1.1	1.3	0	1.2	1.2	0	1.4
	$S_{\cdot 0}(t_1)$	0	0.7		0	0.6		0	0.7
	$S_{\cdot 0}(t_2)$	−0.1	1.8		−0.2	1.5		−0.1	1.7
	$S_{\cdot 0}(t_3)$	−0.1	0.9		−0.1	0.8		−0.1	1

Table 2. Estimates at sample size $N = 800$, 3,200, and 12,800 with 50% censoring: t_1, t_2, and t_3 are the 10%, 50%, and 90%-th percentiles of $f_{\cdot 0}(t)$. Bias is 100× the Monte-Carlo bias, SD* is 100× the Monte-Carlo standard deviation, and RE is the relative efficiency as the ratio of the SD of the full likelihood estimator versus the SD of the other two estimators*

N		Conditional			Weighted			Full	
		Bias*	SD*	RE	Bias*	SD*	RE	Bias*	SD*
800	β_1	−4.6	7.8	0.4	−0.5	6.9	0.5	−1.1	3.2
	$S_{\cdot 0}(t_1)$	−0.1	3.2		−0.9	2.9		0	1.5
	$S_{\cdot 0}(t_2)$	−1.6	8.4		−4	7.8		−0.8	4
	$S_{\cdot 0}(t_3)$	−5.1	6.1		−6.3	5.4		−5.6	5.4
3,200	β_1	−0.5	2.9	0.6	0.4	4.1	0.4	−0.3	1.8
	$S_{\cdot 0}(t_1)$	−0.3	1.6		−0.6	1.6		0	0.8
	$S_{\cdot 0}(t_2)$	−1.6	3.9		−2.5	4.4		−0.3	2.1
	$S_{\cdot 0}(t_3)$	−2.9	2.5		−3.4	2.5		−2.3	2.3
12,800	β_1	−0.2	1.3	0.6	−0.3	2.3	0.4	−0.3	0.8
	$S_{\cdot 0}(t_1)$	−0.1	0.9		−0.1	1		0.1	0.4
	$S_{\cdot 0}(t_2)$	−0.7	2.2		−0.5	2.6		0.2	1
	$S_{\cdot 0}(t_3)$	−1.5	1.4		−1.2	1.5		−0.7	1.2

Table 3. Estimates of β_1 with 50% censoring at a common censoring time. Bias is 100× the Monte-Carlo bias and SD* is 100× the Monte-Carlo standard deviation*

N	Conditional		Weighted		Full	
	Bias*	SD*	Bias*	SD*	Bias*	SD*
1600	−1.5	10.0	1.4	9.1	0.9	9.5
3200	−1.4	6.9	−0.9	6.0	−1.5	6.6
6400	−1.4	4.4	−0.8	4.2	−1.4	4.6
12800	−0.8	2.7	0.3	2.7	−0.3	2.9

5. MRFIT STUDY

5.1 Description

We now apply our method to the MRFIT study to estimate the survival patterns for the compliers and the non-compliers, both under the SI arm and under the UC arm.

Under the SI arm, the compliance groups are identifiable with the passage of time. We let $\mathcal{C}_i = 1$ indicate subject i a complier and $\mathcal{C}_i = 0$ a non-complier. Denote $f^s_{\cdot 0}(f)$ and $f^s_{\cdot 1}(t)$ as the failure densities for the non-compliers and the compliers under SI, then each subject has the failure density

$$\begin{aligned} f^s_i(t) &= (1-\mathcal{C}_i) f^s_{\cdot 0}(t) + \mathcal{C}_i f^s_{\cdot 1}(t), \quad \text{with} \\ f^s_{\cdot 1}(t) &= f^s_{\cdot 0}(t) \exp(\boldsymbol{\beta}'_s \mathbf{t}) \end{aligned} \tag{5}$$

where $\boldsymbol{\beta}_s = (\beta_{s0}, \beta_{s1})'$, $i = 1, \ldots, N_s$, and N_s is the number of participants under SI. Though Kaplan-Meier can estimate the survival functions for the compliers and the non-compliers separately, the estimate for the compliers can be inefficient due to the low compliance rate of 23%.

Under the UC arm, the compliance groups are not identifiable. However, each subject i has a probability p_i to be a complier. Denote $f^u_{\cdot 0}(t)$ and $f^u_{\cdot 1}(t)$ as the failure densities for the non-compliers and the compliers under UC, then each

Table 4. Estimates from MRFIT: slope parameter in the exponential tilting with the standard deviation (SD) and the 95% confidence interval (CI) from 200 bootstraps holding the compliance probability as known

	Weighted likelihood			Full likelihood		
	Estimate	SD	CI	Estimate	SD	CI
β_{s1}	0.28	0.45	$(-0.44,\ 1.19)$	0.31	0.5	$(-0.51,\ 1.43)$
β_{u1}	1.54	0.93	$(-0.19,\ 3.23)$	1.66	1.01	$(-0.22,\ 3.64)$

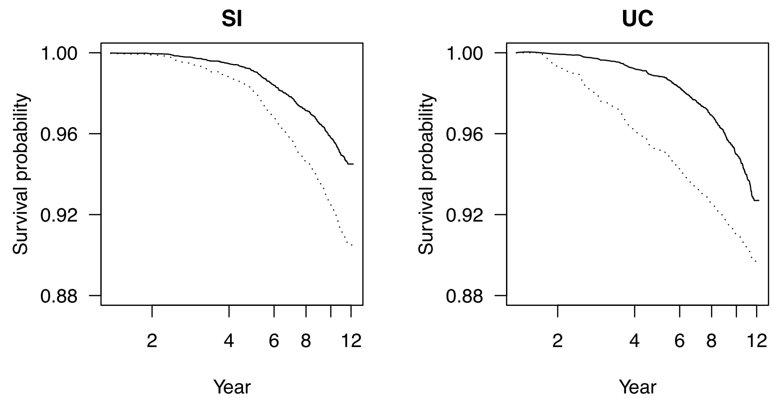

Figure 1. Estimated survival curves from MRFIT: the solid lines for the compliers and the broken lines for the non-compliers.

subject has the failure density

$$
\begin{aligned}
& f_i^u(t) = (1-p_i) f_{.0}^u(t) + p_i f_{.1}^u(t), \quad \text{with} \\
(6) \qquad & f_{.1}^u(t) = f_{.0}^u(t) \exp(\boldsymbol{\beta}_u' \mathbf{t})
\end{aligned}
$$

where $\boldsymbol{\beta}_u = (\beta_{u0}, \beta_{u1})'$, $i = 1, \ldots, N_u$, and N_u is the number of subjects under UC. For model (5) and (6) to better fit the data, we take t as the logarithm of the days post randomization.

5.2 Estimation

The UC arm follows (6), which is the semiparametric mixture survival model studied in section 2. The SI arm follows (5), a special case of the mixture model with $p_i = \mathcal{C}_i$ as 0 or 1. Here, we allow SI and UC to have different exponential tilting parameters for possibly varying compliance effect under the two arms.

Parameter estimates are presented in Table 4. The survival function estimates are presented in Figure 1. In Table 4, we report the slope parameters β_{s1} and β_{u1} from the weighted and the full likelihood estimations. The conditional likelihood estimates are quite close to the weighted likelihood estimates. With the linear exponential tilting, the slope parameters β_{s1} and β_{u1} reflect the differences between the compliance groups: a positive value means the compliers have higher survival than the non-compliers. We see positive slope estimates for both SI and UC. However, the confidence intervals indicate no significant impact of compliance on survival. This lack of significance could be due to the insufficient sample size in the study, whose original goal was not to evaluate the compliance effect. In this example, the full likelihood estimates do not have better efficiency than the weighted likelihood estimates, which may be because all censoring occurred in the last two years of follow up.

To check on the goodness-of-fit of (6), we can add in higher order terms of t in the exponential tilting, and judge by the closeness of the survival curve estimates from a higher order exponential tilting to those from the linear exponential tilting. For the goodness-of-fit of (5), we can compare the semiparametric survival function estimates with the Kaplan-Meier estimates. Our results indicate that (5) and (6) fit the data well.

5.3 Treatment effect

In Table 4, the estimate of β_{u1} is higher than that of β_{s1}. However, it does not mean that the compliers under UC have higher survival than the compliers under SI. In fact, β_{s1} and β_{u1} are not directly comparable as they are tilting from different baseline densities $f_{.0}^s(t)$ and $f_{.0}^u(t)$, respectively.

To evaluate the treatment effect for the compliers, we can directly compare $\widehat{S}_{.1}^s(t)$ with $\widehat{S}_{.1}^u(t)$. Alternatively, we can compare the summary statistics of the survival functions, for example, the restricted mean lifetime $\int_0^\tau \widehat{S}_{.1}^s(t)dt$ versus $\int_0^\tau \widehat{S}_{.1}^u(t)dt$ due to the high percentage of censoring (Irwin, 1949; Zucker, 1998). We take $\tau = 3782$, the day of the first censoring. Table 5 presents the estimates, indicating no significant benefit from SI over the compliers. Since compliance has no significant impact on survival, the estimates over compliers are not very different from the estimates over all participants based on ITT Kaplan-Meier analysis.

Table 5. Estimates from MRFIT: estimates of the restricted mean lifetime (days) with the 95% confidence intervals (CI) from 200 bootstraps holding the compliance probability as known

	SI		UC	
	Estimate	CI	Estimate	CI
Over compliers	3,691	[3,660, 3,716]	3,673	[3,636, 3,714]
Over all participants	3,635	[3,615, 3,652]	3,616	[3,578, 3,638]

6. CONCLUSION

We introduce a semiparametric mixture survival model. It applies when there exist two underlying survival patterns in the population, with membership of each subject to the survival patterns not observable but represented by a probability. The primary interest is the estimation of the survival patterns and whether they are the same.

Three semiparametric likelihood estimations are developed for model estimation. Based on numerical studies, the weighted likelihood estimator can have better efficiency than the conditional likelihood estimator when the sample size is moderate and the estimated censoring probability is not low at the uncensored observations, but the latter can be more efficient otherwise. The full likelihood estimator resembles the "re-distribution to the right" algorithm but with differential re-distribution weights. The improvement in efficiency from full likelihood estimation depends on the pattern of censoring.

The efficiency of the semiparametric mixture model is fundamentally affected by the membership probability $\{p_i, i = 1, \ldots, N\}$. The semiparametric estimators have better efficiency if p_i's are more symmetrically distributed, for example, $p_i \sim Beta(1, 1)$ versus $p_i \sim Beta(2, 1/2)$. The semiparametric estimators also have better efficiency if p_i's are more densely distributed around 0 and 1. Two extreme cases for the last point are: (1) if the membership is totally known, $p_i = 0$ or 1, the semiparametric estimates have better efficiency than with p_i in $(0, 1)$; (2) if p_i's all around 1/2, the semi parametric estimators are unstable. These extreme scenarios are observable in Table 4 for the MRFIT study, where the estimates are less variable with $p_i = 0$ or 1 in SI arm than p_i around 0.5 in UC arm.

Though this work treats the membership as known, it applies when the membership is unknown but consistently estimated.

APPENDIX A. TWO-STEP ITERATIVE EM ALGORITHM FOR FULL LIKELIHOOD ESTIMATION

We use a two-step iterative EM algorithm for the estimation of $(\boldsymbol{\beta}, \mathbf{q}_0)$ from (4). We first pick initial values $q_{0j}^{(0)}$ for $f_{\cdot 0}(t_j)$ and $q_{1j}^{(0)}$ for $f_{\cdot 1}(t_j)$. In step 1, we compute $w_{ij}^{(0)}$ by

$$w_{ij}^{(0)} = \delta_i I(x_i = t_j) + (1 - \delta_i) \frac{\{(1 - p_i) q_{0j}^{(0)} + p_i q_{1j}^{(0)}\} I(t_j \geq x_i)}{\sum_{l=1}^{n} \{(1 - p_i) q_{0l}^{(0)} + p_i q_{1l}^{(0)}\} I(t_l \geq x_i)}$$

for $i = 1, \ldots, N$, $j = 1, \ldots, n$. In step two, we let $w_{ij} = w_{ij}^{(0)}$ be fixed and estimate $(\boldsymbol{\beta}, \mathbf{q}_0)$ as the maximizer of (4). Following similar derivations as in the previous two sections, the estimate of q_{0j} at fixed $\boldsymbol{\beta}$ has the form

$$q_{0j} = \frac{1}{N} \left(\sum_{i=1}^{N} w_{ij} \right) \frac{1}{1 + \rho\{\exp(\boldsymbol{\beta}' \mathbf{t}_j) - 1\}}$$

with

$$\rho = \rho(\boldsymbol{\beta}) = N^{-1} \sum_{j=1}^{n} \sum_{i=1}^{N} w_{ij} \frac{p_i \exp(\boldsymbol{\beta}' \mathbf{t}_j)}{(1 - p_i) + p_i \exp(\boldsymbol{\beta}' \mathbf{t}_j)}.$$

The profile log-likelihood of $\boldsymbol{\beta}$ is then

$$l_e(\boldsymbol{\beta}) = \sum_{j=1}^{n} \sum_{i=1}^{N} w_{ij} \big(\log\{(1 - p_i) + p_i \exp(\boldsymbol{\beta}' \mathbf{t}_j)\} - \log[1 + \rho\{\exp(\boldsymbol{\beta}' \mathbf{t}_j) - 1\}] \big).$$

Let $\boldsymbol{\beta}^{(1)}$ be the maximizer of $l_e(\boldsymbol{\beta})$. We then update q_{0j} and q_{1j} at $\boldsymbol{\beta}^{(1)}$ and go back to step 1. The two steps are repeated until convergence.

APPENDIX B. CONSISTENCY OF SEMIPARAMETRIC LIKELIHOOD ESTIMATOR

Consistency is developed under the regularity conditions: (1) the underlying survival functions $S_{\cdot 0}$ and $S_{\cdot 1}$ are non-degenerate; (2) p_i's are identical and independently distributed over $[0, 1]$; (3) Let t_n be the last observed event time, $t_n \to \infty$ as $N \to \infty$.

Under the regularity conditions, $\boldsymbol{\beta}$ is root-n consistent with $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N(0, U)$, and the survival function estimate $\widehat{S}_{\cdot 0}(t)$ satisfies $\sqrt{n}\{\widehat{S}_{\cdot 0}(t) - S_{\cdot 0}(t)\} \to B_{\cdot 0}(t)$ with $B_{\cdot 0}(t)$ a mean zero Gaussian process.

We give a sketch of the proof taking the weighted likelihood estimator as an example. Denote the true value of $\boldsymbol{\beta}$ as $\boldsymbol{\beta}_T$. The estimate $\widehat{\boldsymbol{\beta}}$ is maximizer of $l_{pw}(\boldsymbol{\beta})$ with ρ satisfying

$$\text{(B.1)} \qquad \sum_{i=1}^{n} 1/g_i \frac{w(t_i;\boldsymbol{\beta})}{1+\rho w(t_i;\boldsymbol{\beta})} = 0$$

where $w(t;\boldsymbol{\beta}) = \exp(\boldsymbol{\beta}'\mathbf{t}) - 1$. The profile likelihood l_{pw} can be written as $l_{pw}(\theta,\boldsymbol{\beta}) = l_1 - l_2$ with

$$l_1 = \sum_{i=1}^{n} 1/g_i \log\{(1-p_i) + p_i \exp(\boldsymbol{\beta}'\mathbf{t}_i)\},$$
$$l_2 = \sum_{i=1}^{n} 1/g_i \log\{1 + \rho w(t_i;\boldsymbol{\beta})\}.$$

The estimate $\widehat{\boldsymbol{\beta}}$ is solution to

$$\frac{\partial l_{pw}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial l_1}{\partial \boldsymbol{\beta}} - \frac{\partial l_2}{\partial \boldsymbol{\beta}} = 0,$$

where due to (B.1), we have

$$\partial l_2/\partial \boldsymbol{\beta} = \rho \sum_{i=1}^{n} 1/g_i \frac{\partial w(t_i;\boldsymbol{\beta})/\partial \boldsymbol{\beta}}{1+\rho w(t_i;\boldsymbol{\beta})}.$$

Expand $\partial l_{pw}(\widehat{\boldsymbol{\beta}})/\partial\beta$ at $\boldsymbol{\beta}_T$, we see that

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_T = A_n^{-1} b_n + o_p(n^{-1/2}),$$

where

$$A_n = -\frac{1}{n}\frac{\partial^2 l_{pw}(\boldsymbol{\beta}_T)}{\partial^2 \boldsymbol{\beta}}, \quad b_n = \frac{1}{n}\frac{\partial l_{pw}(\boldsymbol{\beta}_T)}{\partial \boldsymbol{\beta}}.$$

We can show that $A_n \to A$ in probability for some symmetric matrix A and $\sqrt{n} b_n \to N(0,V)$ for some semi-positive definite matrix V. The asymptotic normality of $\widehat{\boldsymbol{\beta}}$ follows with $U = A^{-1}VA^{-1}$.

Consistency of $\widehat{S}_{\cdot 0}$ can be similarly proved as in Qin (1999).

Received 11 December 2012

REFERENCES

Anderson, J. A. (1979). Multivariate logistic compounds. *Biometrika* **66** 17–26. MR0529143

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B* **39** 1–38. MR0501537

Efron, B. (1967). The two sample problem with censored data. *Proceedings of the Fifth Berkley Symposium on Mathematical Statistics and Probability* **4** 831–853.

Efron, B. and Feldman, D. (1991). Compliance as an explanatory variable in clinical trials. *Journal of the American Statistical Association* **86** 9–18.

Follmann, D. A. (2000). On the effect of treatment among would-be treatment compliers: An analysis of the multiple risk factor intervention trial. *Journal of the American Statistical Association* **95** 1101–1109. MR1821718

Hall, P. and Titterington, D. M. (1984). Efficient nonparametric estimation of mixture proportions. *Journal of the Royal Statistical Society: Series B* **46** 465–473. MR0790632

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47** 663–685. MR0053460

Irwin, J. O. (1949). The standard error of an estimate of expectational life. *Journal of Hygiene* **47** 188–189.

Ma, Y., Hart, J. D. and Carroll, R. J. (2011). Density estimation in several populations with uncertain population membership. *Journal of the American Statistical Association* **106** 1180–1192. MR2894773

McLachlan, G. J. and Basford, K. E. (1988). *Mixture models. Inference and application in clustering.* New York: Marcel Dekker. MR0926484

Multiple-Risk-Factor-Intervention-Trial-Group (1982). Multiple risk factor intervention trial risk factor changes and mortality results. *Journal of the American Medical Association* **248** 1465–1477.

Multiple-Risk-Factor-Intervention-Trial-Group (1996). Mortality after 16 years for participants randomized to multiple risk factor intervention trial. *Circulation* **94** 946–951.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249. MR0946049

Qin, J. (1999). Empirical likelihood ratio based confidence intervals for mixture proportions. *Annals of Statistics* **27** 1368–1384. MR1740107

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics* **22** 300–325. MR1272085

Qin, J. and Leung, H. Y. (2005). A semiparametric two-component "compound" mixture model and its application to estimating malaria attribution fractions. *Biometrics* **61** 456–464. MR2140917

Wang, C. Y., Wang, S., Gutierrez, R. G. and Carroll, R. J. (1998). Local linear regression for generalized linear models with missing data. *Annals of Statistics* **26** 1028–1050. MR1635438

Zucker, D. M. (1998). Restricted mean life with covariates: modification and extension of a useful survival analysis method. *Journal of the American Statistical Association* **93** 702–709. MR1631365

Zonghui Hu
Biostatistics Research Branch
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, MD 20892-7609
USA
E-mail address: huzo@niaid.nih.gov

Jing Qin
Biostatistics Research Branch
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, MD 20892-7609
USA
E-mail address: jqin@niaid.nih.gov

Dean A. Follmann
Biostatistics Research Branch
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, MD 20892-7609
USA
E-mail address: dfollmann@niaid.nih.gov