STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:3089–3110
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2174

Adjusted Kaplan–Meier estimator and log-rank test with inverse probability of treatment weighting for survival data

Jun Xie[1,*,†] and Chaofeng Liu[2]

[1]*Department of Statistics, Purdue University, 150 N. University Street, West Lafayette, IN 47907-2067, U.S.A.*
[2]*Lilly Research Laboratory, Indianapolis, IN 46285, U.S.A.*

SUMMARY

Estimation and group comparison of survival curves are two very common issues in survival analysis. In practice, the Kaplan–Meier estimates of survival functions may be biased due to unbalanced distribution of confounders. Here we develop an adjusted Kaplan–Meier estimator (AKME) to reduce confounding effects using inverse probability of treatment weighting (IPTW). Each observation is weighted by its inverse probability of being in a certain group. The AKME is shown to be a consistent estimate of the survival function, and the variance of the AKME is derived. A weighted log-rank test is proposed for comparing group differences of survival functions. Simulation studies are used to illustrate the performance of AKME and the weighted log-rank test. The method proposed here outperforms the Kaplan–Meier estimate, and it does better than or as well as other estimators based on stratification. The AKME and the weighted log-rank test are applied to two real examples: one is the study of times to reinfection of sexually transmitted diseases, and the other is the primary biliary cirrhosis (PBC) study. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: adjusted Kaplan–Meier estimator; IPTW; survival function; weighted log-rank test

1. INTRODUCTION

Kaplan–Meier estimates of survival functions are widely used in many clinical studies. In a randomized trial, patients are randomly assigned to different groups with no systematic difference between covariate factors. In this situation, the Kaplan–Meier estimators and their statistical comparison with the log-rank test are usually adequate. However, in a non-randomized clinical trial or an observational study, the samples in different groups may be biased due to some confounding variables, hence the Kaplan–Meier estimator is inappropriate.

When confounders are present, survival function estimates can be adjusted and compared using specified models such as the proportional hazards model [1] and the additive risk model [2]. The calculation of expected survival curves for subjects in a study group, based

*Correspondence to: Jun Xie, Department of Statistics, Purdue University, 150 N. University Street, West Lafayette, IN 47907-2067, U.S.A.
†E-mail: junxie@stat.purdue.edu

Received 4 January 2004
Accepted 21 December 2004

on the proportional hazards model of their observed covariates, has been discussed by some researchers [3, 4]. These methods allow for adjustment of the effects of other covariates, but in some situations, the proportional or additive assumption is invalid. Other adjustments of survival estimation based on matching or stratification have been proposed [5–8], which stratify observations according to the values of some confounding variables and then combine the survival estimates in each stratum. These estimators require observations available at time t in all the strata. When some strata have small numbers of individuals, the survival function can only be estimated for a small part of the observation time. Winnett and Sasieni [9] developed a weighted Nelson–Aalen estimator more suitable for smaller strata. The weighted Kaplan–Meier estimator for stratified data was further applied by Galimberti *et al.* [10]. However, stratification and matching approaches may have difficulties to get well-matched data when a confounding variable is continuous or many confounding variables have to be considered.

In this article, we propose a new approach to reduce confounding effects. We aim at estimating the marginal survival curves but do not use stratification. Specifically, we develop an adjusted Kaplan–Meier estimator (AKME) using inverse probability of treatment weighting (IPTW). A weight is assigned to each subject as the inverse probability of being in a certain group conditioning on a set of covariates. If a subject has a higher probability of being in a group, it is considered as over-represented, and therefore is given a lower weight. On the other hand, if the subject has a smaller probability of being in the group, it is considered as under-represented and is given a higher weight. Generally, in an observational study, the weighting adjustment will remove sampling bias. Even in a randomized trial for a treatment, other interesting factors, such as gender or different age levels may not be randomized. When we study the survival curves of those groups, the AKME will provide comparable estimates.

The IPTW method [11] has been applied in many research fields such as design and analysis of two-stage studies [12], regression analysis with missing covariate data [13], estimating effects of time-varying treatments on the discrete-time hazard [14], and estimation of casual treatment effects [15]. Particularly, these weighting approaches are related to the concept of propensity score, which is defined by Rosenbaum and Rubin [16] as the conditional probability of assignment to a certain group given a vector of observed covariates. Applying the IPTW method in survival estimation, we adjust for confounding by using estimated propensity scores to construct weights for individual observations.

In Section 2, the AKME is developed, with discussion of its application to estimating survival functions of different groups. In Section 3, a pseudo-likelihood function is defined and the AKME is shown to be the maximum pseudo-likelihood estimate. The asymptotic mean and variance are also derived. In Section 4, a weighted log-rank test is proposed for comparing survival functions of different groups. Under the null hypothesis of no difference between two groups, the weighted log-rank statistic has an asymptotic standard normal distribution. For small-to-moderate sample size, a bootstrap approach is developed. In Section 5, we use simulation studies to illustrate the estimation procedure and evaluate the performance of hypothesis tests. Applications to two clinical studies are presented in Section 6.

2. THE ADJUSTED KAPLAN–MEIER ESTIMATOR

Let $(T_i, \delta_i, X_i, \mathbf{Z}_i)$, $i = 1, \ldots, n$, denote an independent sample of right-censored survival data with two or more groups, where T_i is the possibly right-censored event time, δ_i is the censoring

indicator, $\delta_i = 0$ if T_i is censored and $\delta_i = 1$ if T_i corresponds to an event; X_i is the group index, $X_i = 1, \ldots, K$ for K different groups, and $\mathbf{Z}_i$ is the covariate vector.

Let p_{ik} be the probability of the ith individual being in group k. This probability may depend on the covariate vector $\mathbf{Z}_i$, i.e. $p_{ik} = P(X_i = k|\mathbf{Z}_i)$. If certain covariate values that correspond to low survival rates are more strongly represented in one group than another, then survival estimated by the Kaplan–Meier method from one group would appear to be worse than survival estimated from the other group. Therefore, the group effect represented by those subjects in group k is confounded by the covariates. Let $S^k(t)$ denote the survival function of group k. We develop an AKME of the survival function of each group so that the groups can be compared. In this article, we assume that (a) p_{ik}'s are either known (as in a designed study) or can be consistently estimated given $\mathbf{Z}_i$, and (b) p_{ik}'s are bounded away from 0.

Suppose the events (further also referred to as deaths) occur at D distinct times $t_1 < t_2 < \cdots < t_D$ in the whole sample. Ties are allowed. At time t_j, $j = 1, \ldots, D$, there are d_{jk} deaths out of Y_{jk} individuals at risk in group k. Then we can write $d_{jk} = \sum_{i:T_i = t_j} \delta_i I(X_i = k)$ and $Y_{jk} = \sum_{i:T_i \geqslant t_j} I(X_i = k)$, where $I(X_i = k)$ is the indicator function that equals 1 if subject i belongs to group k and 0 otherwise.

To reduce the sample bias of different groups, we assign a weight $w_{ik} = 1/p_{ik}$ for the ith individual, when it is in the kth group, $i = 1, \ldots, n$. Then the weighted number of events and the weighted number at risk in group k are defined as

$$d_{jk}^w = \sum_{i:T_i = t_j} w_{ik} \delta_i I(X_i = k) = \sum_{i:T_i = t_j} \frac{\delta_i I(X_i = k)}{p_{ik}}$$

$$Y_{jk}^w = \sum_{i:T_i \geqslant t_j} w_{ik} I(X_i = k) = \sum_{i:T_i \geqslant t_j} \frac{I(X_i = k)}{p_{ik}}$$

The following formula defines the AKME for the kth group:

$$\hat{S}^k(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_j \leqslant t} [1 - d_{jk}^w / Y_{jk}^w] & \text{if } t_1 \leqslant t \end{cases} \tag{1}$$

if $Y_{jk}^w > 0$. As usual, the AKME should be done in practice only up to times at which individuals are at risk. Notice that in randomized clinical trials, subjects would have equal probabilities of being in each group. Consequently, the AKME is reduced to the Kaplan–Meier estimator.

3. PROPERTY OF THE ESTIMATOR

From now on, we consider the case with only two groups, denoted by $X = 0$ and $X = 1$. We refer to group 0 as the control and group 1 as the treatment group. In practical situations, the two groups may be defined by other features, such as male and female. Let $p_i = P(X_i = 1|\mathbf{Z}_i)$. Then $1 - p_i = P(X_i = 0|\mathbf{Z}_i)$. The weight for an individual in the treatment group becomes $1/p_i$ and the weight for an individual in the control group is $1/(1 - p_i)$.

3.1. Pseudo-likelihood and the AKME

The Kaplan–Meier estimate was shown to be the non-parametric maximum-likelihood estimate [17] under appropriate conditions. Here, we define a pseudo-likelihood function for survival data and show that AKME is the maximum pseudo-likelihood estimate. A general notation of the survival function $S(t)$ is used, where

$$S(t) = X \cdot S^1(t) + (1 - X) \cdot S^0(t)$$

This survival function is a random quantity whose value depends on the value of X. More specifically, for the sample $i = 1, \ldots, n$, we write $S_i(t) = X_i \cdot S^1(t) + (1 - X_i) \cdot S^0(t)$.

Consider the observations in one group, for example the treatment group. A pseudo-likelihood function is defined as

$$\prod_{i=1}^{n} ([S_i(T_i - 0) - S_i(T_i)]^{\delta_i} [S_i(T_i)]^{1-\delta_i})^{X_i/p_i}$$

where $S_i(T_i - 0)$ includes but $S_i(T_i)$ excludes the probability of death exactly at T_i. The log-pseudo-likelihood is written as

$$\sum_{i=1}^{n} \frac{X_i}{p_i} [\delta_i \log(S_i(T_i - 0) - S_i(T_i)) + (1 - \delta_i) \log(S_i(T_i))]$$

This likelihood is available only for the individuals in the treatment group where $X_i = 1$.

A similar definition has been used by other researchers in multi-state models and two-stage studies [12, 18, 19]. Introducing the pseudo-likelihood to the survival function, we can show that if the treatment assignment X_i is independent of both censoring and the survival processes conditioning on the covariate vector $\mathbf{Z}_i$, then

$$E\left\{\sum_{i=1}^{n} \frac{X_i}{p_i} [\delta_i \log(S_i(T_i - 0) - S_i(T_i)) + (1 - \delta_i) \log(S_i(T_i))]\right\}$$
$$= E\left\{\log\left(\prod_{i=1}^{m} [S^1(T_i - 0) - S^1(T_i)]^{\delta_i} [S^1(T_i)]^{1-\delta_i}\right)\right\} \quad (2)$$

where $m = \sum_{i=1}^{n} X_i$ is the number of subjects in the treatment group. Equation (2) links the pseudo-likelihood to the likelihood. We give a detailed derivation in Appendix A.1. The following proposition shows that the AKME is an optimal estimator.

Proposition 1
The AKME defined in Formula (1) maximizes the pseudo-likelihood if the probability p_i is known.

The proof is provided in Appendix A.2.

3.2. Mean and variance estimations

To derive the mean and variance estimations of AKME, we use some intermediate results from the proof of Proposition 1. More specifically, in Appendix A.2, we defined $s_j^1 = S^1(t_j)/S^1(t_{j-1})$ and its adjusted estimator $\hat{s}_j^1 = 1 - d_{j1}^w / Y_{j1}^w$, where $t_1 < t_2 < \cdots < t_k$ denote the distinct times of death in the set of observed T_i in the treatment group. Let E_j denote a conditional expectation

given information up to time t_j. If the number of individuals at risk is positive, i.e. $Y_{j1}>0$, then we obtain $E_j(\hat{s}_j^1)=s_j^1$ as shown below:

$$\begin{aligned}
E_j(1-\hat{s}_j^1) &= E_j\left[\frac{\sum_{i:T_i=t_j}\frac{X_i}{p_i}\cdot\delta_i}{\sum_{i:T_i\geqslant t_j}\frac{X_i}{p_i}}\right] \\
&= E_j\left\{E_j\left[\frac{\sum_{i:T_i=t_j}\frac{X_i}{p_i}\cdot\delta_i}{\sum_{i:T_i\geqslant t_j}\frac{X_i}{p_i}}\middle| X_i,\mathbf{Z}_i,i \text{ s.t. } T_i\geqslant t_j\right]\right\} \\
&= E_j\left\{\frac{(1-s_j^1)\sum_{i:T_i\geqslant t_j}\frac{X_i}{p_i}}{\sum_{i:T_i\geqslant t_j}\frac{X_i}{p_i}}\right\} \\
&= 1-s_j^1 \qquad (3)
\end{aligned}$$

Let $T_{\max}$ be the largest observed time in the treatment group, i.e. $T_{\max}=\max\{T_i, 1\leqslant i\leqslant n \text{ and } X_i=1\}$. Given time $t\in[0,T_{\max})$, let t_l be the largest observed death time satisfying $t_l\leqslant t$. Thus, $t_l\leqslant t<T_{\max}$ and we have $Y_{l1}>0$. By means of successive conditional expectations, we obtain

$$\begin{aligned}
E[\hat{S}^1(t)] &= E[\hat{s}_1^1\cdots\hat{s}_{l-1}^1\cdot E_l(\hat{s}_l^1)] = E[\hat{s}_1^1\cdots\hat{s}_{l-1}^1 s_l^1] \\
&= s_l^1 E[\hat{s}_1^1\cdots\hat{s}_{l-2}^1\cdot E_{l-1}(\hat{s}_{l-1}^1)] \\
&= \cdots = s_{1\ldots}^1 s_{l-1}^1 s_l^1 = S^1(t)
\end{aligned}$$

Therefore, in the range where we have data, the AKME is an unbiased estimator. However, for time t to the right of $T_{\max}$, the AKME is biased upward [20]. If $T_{\max}$ corresponds to a death time, then the AKME of the survival curve is zero beyond this point. If $T_{\max}$ is censored, the value of $S(t)$ beyond this point is undetermined because we do not know when this last survivor would have died if the survivor had not been censored. In practice, the AKME of the survival function is well defined for all time points less than $T_{\max}$.

Proposition 2
Assume $\max_{i:T_i\geqslant t_j}(1/p_i)/\sum_{i:T_i\geqslant t_j}1/p_i \underset{n\to\infty}{\to} 0$.

(a) If the probabilities of treatment p_i's are known, the variance of the AKME is given by

$$\mathrm{Var}[\hat{S}^1(t)] = (S^1(t))^2\sum_{j:t_j\leqslant t}\frac{1-s_j^1}{M_j s_j^1} \qquad (4)$$

where $M_j=(\sum_{i:T_i\geqslant t_j}1/p_i)^2/\sum_{i:T_i\geqslant t_j}(1/p_i)^2$.

(b) If the probabilities p_i's are unknown but estimated given the covariate vector $\mathbf{Z}_i$ (by a parametric model, or a non-parametric smoothing method [13]), then the variance of the AKME is estimated by

$$\hat{V}[\hat{S}^1(t)] = [\hat{S}^1(t)]^2\sum_{j:t_j\leqslant t}\frac{1-\hat{s}_j^1}{\hat{M}_j\hat{s}_j^1} \qquad (5)$$

where $\hat{S}^1(t)$ and $\hat{s}_j^1$ are the AKME, and $\hat{M}_j$ is calculated at $\hat{p}_i$, $i=1,\ldots,n$.

The derivation of the variance formula is provided in Appendix A.3.

If the probabilities of being in the treatment, p_i's, are the same, then the variance estimation, Formula (5), reduces to Greenwood's formula of the variance of the Kaplan–Meier estimator. Under the condition that the probability of being in the treatment group is bounded away from zero, that is, there exists a positive value ε such that $p_i \geqslant \varepsilon > 0$ for all $i = 1, \ldots, n$, $\mathrm{Var}[\hat{S}^1(t)]$ would have the same rate of convergence to zero as the variance of the Kaplan–Meier estimator. Therefore, the AKME is a consistent estimator.

Next, consider a simple example with two groups, treatment and control. The hazard rate functions for the treatment and control groups are the same with

$$h^1(t,z) = h^0(t,z) = \begin{cases} 0.5, & z = 1 \\ 2.5, & z = 0 \end{cases}$$

where the hazard function depends on a binary covariate z, with $z \sim \text{Bernoulli}(0.5)$. The true survival function, integrating out z, is $S^1(t) = S^0(t) = 0.5\mathrm{e}^{-0.5t} + 0.5\mathrm{e}^{-2.5t}$.

Suppose the probability of being in the treatment group depends on the covariate z:

$$P(X = 1|z) = \begin{cases} 0.75, & z = 1 \\ 0.25, & z = 0 \end{cases}$$

It is easy to calculate that $P(z = 1|X = 1) = 0.75$. Therefore, the Kaplan–Meier estimators of the treatment group converge to $\tilde{S}^1(t) = 0.75\mathrm{e}^{-0.5t} + 0.25\mathrm{e}^{-2.5t}$, which is the marginal survival function of the biased sample in the treatment group. Similarly, we have $P(z = 1|X = 0) = 0.25$; hence the limit of the Kaplan–Meier estimator of the control group is $\tilde{S}^0(t) = 0.25\mathrm{e}^{-0.5t} + 0.75\mathrm{e}^{-2.5t}$.

In the AKME, individual i is assigned the weight $1/P(X_i = 1|Z_i)$ if it is in the treatment group, or $1/(1 - P(X_i = 1|Z_i))$ if it is in the control group. Therefore, the limits of the AKME would be $S^1(t) = S^0(t) = 0.5\mathrm{e}^{-2.5t} + 0.5\mathrm{e}^{-0.5t}$, which are the true survival functions. Figure 1 shows the different limits of survival curves by the AKME and Kaplan–Meier estimator.

4. HYPOTHESIS TESTING

Consider the hypotheses

$$H_0 : S^1(t) = S^0(t) \quad \text{for all } t \leqslant \tau, \ \textit{versus}$$

$$H_A : S^1(t) \neq S^0(t) \quad \text{for some } t \leqslant \tau$$

where τ is the largest time at which both groups have at least one subject at risk. Let $d_j = d_{j1} + d_{j0}$ and $Y_j = Y_{j1} + Y_{j0}$ be the number of deaths and the number at risk in the whole sample at time t_j, $j = 1, \ldots, D$. The standard log-rank test [21] is based on the statistic

$$G = \sum_{j=1}^{D} \left(d_{j1} - Y_{j1} \left(\frac{d_j}{Y_j} \right) \right)$$

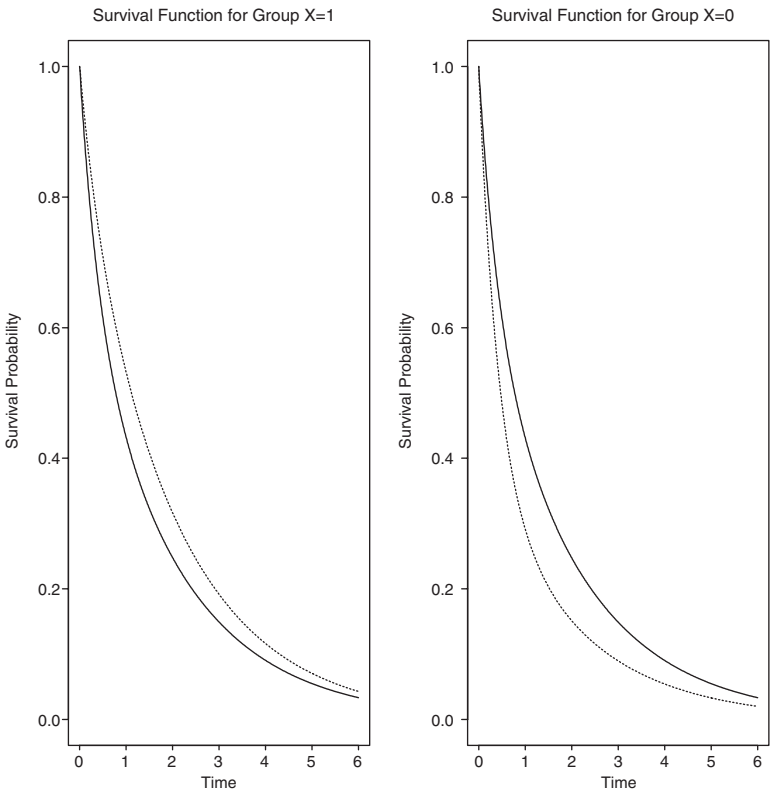

Figure 1. Comparisons of the limits of the AKME and the Kaplan–Meier estimators, where the limit of AKME is the right survival function. The limit of AKME (——); the limit of Kaplan–Meier (……).

with variance

$$\mathrm{Var}(G) = \sum_{j=1}^{D} \frac{Y_{j1}}{Y_j}\left(1 - \frac{Y_{j1}}{Y_j}\right)\left(\frac{Y_j - d_j}{Y_j - 1}\right) d_j$$

However, when the samples in different groups are biased with confounding variables, we use the AKME to estimate survival functions and we propose a weighted log-rank test for statistical comparison of survival functions of the treatment and control groups.

To combine the weighted number of events and the weighted number of individuals at risk into a pooled sample, we need to adjust the weights in each group. At time t_j, $j = 1, \ldots, D$, the weight is reassigned as $Y_{j1} \cdot w_i / \sum_{i:X_i=1} w_i$, for individual i in the treatment group, and $Y_{j0} \cdot w_i / \sum_{i:X_i=0} w_i$, for individual i in the control group. Therefore, the weights are proportional to the number of individuals at risk in each group. Notice that the AKME of the survival curves do not change with the reweighting approach, because the weight coefficient $Y_{j1} / \sum_{i:X_i=1} w_i$ is cancelled in the numerator and the denominator in Formula (1).

Let $d_j^w = d_{j1}^w + d_{j0}^w$ and $Y_j^w = Y_{j1}^w + Y_{j0}^w$ denote the weighted number of deaths and the weighted number at risk in the combined sample at time t_j. The test of H_0 is based on the statistic

$$G^w = \sum_{j=1}^{D} \left(d_{j1}^w - Y_{j1}^w \left(\frac{d_j^w}{Y_j^w} \right) \right)$$

which is a weighted form of the standard log-rank test statistic. The variance of G^w is given by

$$\mathrm{Var}(G^w) = \sum_{j=1}^{D} \left\{ \frac{d_j(Y_j - d_j)}{Y_j(Y_j - 1)} \sum_{i=1}^{Y_j} \left[\left(\frac{Y_{j0}^w}{Y_j^w} \right)^2 w_i^2 X_i + \left(\frac{Y_{j1}^w}{Y_j^w} \right)^2 w_i^2 (1 - X_i) \right] \right\}$$

It is derived by moment formulas of the sum of a random sample drawn from zero-sum scores [24]. The details are described in Appendix A.4.

The weighted log-rank test statistic is proposed as $Z_0 = G^w / \sqrt{\mathrm{Var}(G^w)}$. The test statistic has a standard normal distribution for large samples under the null hypothesis H_0. Thus, the null hypothesis is rejected when $|Z_0| > Z_{\alpha/2}$ at an α level, where $Z_{\alpha/2}$ is the critical value from the standard normal distribution.

Alternatively, for data with small or moderate sample size, we propose a bootstrap approach to perform the hypothesis test as follows: given the estimated probability of treatment $\hat{p}_i$, $i = 1, \ldots, n$, each individual is reassigned to the treatment or control group according to the random variable $X_i \sim \mathrm{Bernoulli}(\hat{p}_i)$. This procedure produces a sample of size n under H_0, because the treatment and control survival times are interchangeable. Repeat the resampling procedure B times, for example $B = 1000$ times, and obtain the weighted log-rank statistic Z_b in the bth simulation, $b = 1, \ldots, B$. Then the bootstrap p-value associated with the test statistic Z_0 is calculated as $p = \sum_b I(|Z_b| \geqslant |Z_0|)/B$.

5. SIMULATION STUDIES

5.1. AKME for data from a proportional hazards model

A sample of size 400 is generated as follows: (1) the covariate $z_i, i = 1, \ldots, 400$, are generated from a Bernoulli distribution with $z \sim \mathrm{Bernoulli}(0.5)$; (2) the group indicator variable $x_i, i = 1, \ldots, 400$ are generated with the probability

$$P(x = 1|z) = \begin{cases} 0.75, & z = 1 \\ 0.25, & z = 0 \end{cases}$$

(3) the survival times are simulated from a proportional hazards function $h(t, x, z) = h_0(t) \exp(\alpha z + \beta x)$, where $h_0(t) = 2.5$, $\alpha = -\ln(5)$, $\beta = 0$; (4) the censoring times are generated according to the exponential distribution with mean γ and an upper limit value of 4, where γ ranges from 1.0 to 3.5 for different censoring rates.

To calculate the AKME, we first estimate the conditional probability $\hat{p}_i$ of being in the treatment group given z_i using logistic regression. Then, the ith subject is assigned a weight $w_i = 1/\hat{p}_i$ if $x_i = 1$ or $w_i = 1/(1 - \hat{p}_i)$ if $x_i = 0$. Using Formula (1), we obtain the AKME for

the treatment and control groups. The regular Kaplan–Meier estimators are also calculated. The AKME is much closer to the true survival curve than the Kaplan–Meier estimator (figure not shown). We could fit the Cox proportional model with the explanatory variable z_i and calculate the expected survival curve [3] for subjects in one group based on their observed covariates. Since proportional hazards model is the true model, the expected survival curve and the AKME curve would both converge to the true survival function.

Alternatively, a stratification method can be used to adjust for confounding variables. In this simple example, there are two strata with $z=1$ and $z=0$, respectively. A weighted Kaplan–Meier estimator for matched data by Galimberti *et al.* [10] is obtained by weighting the number of events and the number of subjects at risk in each stratum by the reciprocal of the stratum size. It can be shown that if the stratum sizes in the whole sample are about the same, as indicated by $z \sim \text{Bernoulli}(0.5)$ in this example, then the weighted Kaplan–Meier estimator from stratification is very close to the AKME. More specifically, consider the survival estimator in the treatment group where $x=1$. In calculating the weighted Kaplan–Meier estimator for the stratified data, the sample size in the stratum with $z=0$ is proportional to $P(z=0|x=1)=0.25$ if there is no sampling variation in the data; thus the weight of this stratum is roughly proportional to $1/0.25$. Similarly, we obtain $P(z=1|x=1)=0.75$ and the weight for the stratum with $z=1$ is roughly proportional to $1/0.75$. On the other hand, in the AKME, all subjects with $z=0$ have the same weight around $1/P(x=1|z=0)=1/0.25$ and all subjects with $z=1$ have the same weight around $1/P(x=1|z=1)=1/0.75$. Therefore, the AKME is very close to the weighted Kaplan–Meier estimator based on stratification.

Next, 15 samples of size 200 are generated from the same model described previously. Figure 2 shows that fifteen AKME curves centre at the true survival curve determined by the model. Standard deviation of the AKME is obtained by Formula (5). In Figure 3, this variance estimate is compared with the Monte Carlo simulated reference, which is obtained as the standard deviation of the AKME calculated from 1000 samples. The curves in Figure 3 give the mean of 1000 standard deviations estimated by Formula (5) and the Monte Carlo reference. It shows that Formula (5) provides good estimates, reasonably close to the Monte Carlo results, with both moderate (25 per cent) and high censoring rate (45 per cent).

5.2. *Hypothesis testing*

The performance of several hypothesis tests is evaluated by simulation studies. We rewrite the parameter as $\beta=\ln(\theta)$ in the previous proportional hazards function $h(t,x,z)=h_0(t)\exp(\alpha z+\beta x)$. Then with $h_0(t)=2.5$ and $\alpha=-\ln(5)$, we get

$$h(t,x,z)=\begin{cases} 0.5, & x=0,\ z=1 \\ 0.5\theta, & x=1,\ z=1 \\ 2.5, & x=0,\ z=0 \\ 2.5\theta, & x=1,\ z=0 \end{cases}$$

where $\theta=1,\ 1.25$, and 1.5 represent the treatment effect with 0, 25, and 50 per cent inflation, respectively, on the hazard rate for the treatment group.

Under each combination of treatment effect and censoring rate, 1000 samples of size 200 are generated and five different test statistics are calculated. The results of hypothesis tests

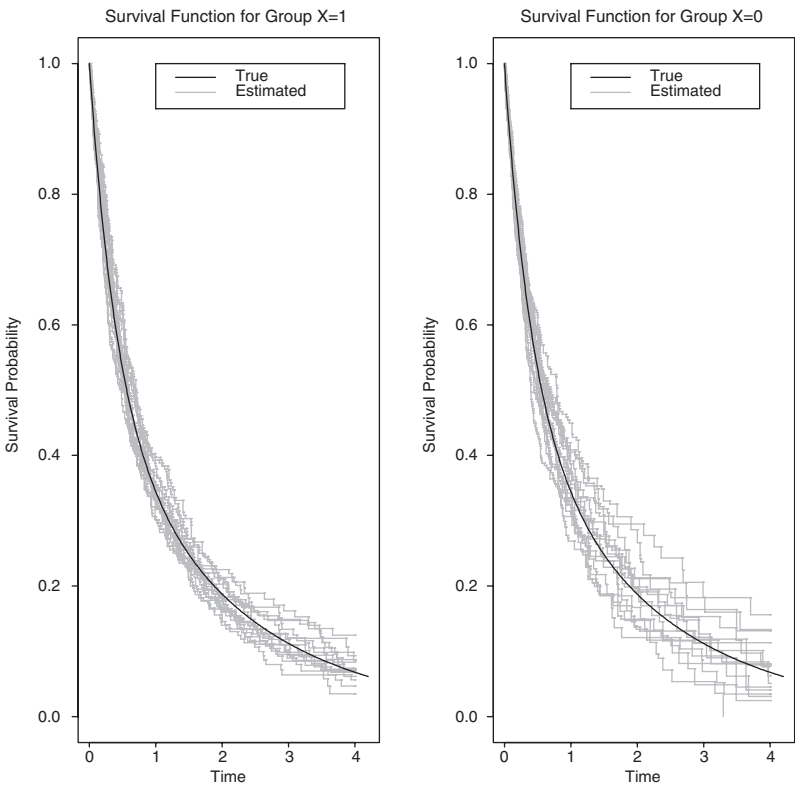

Figure 2. AKME in repeated samples. The AKME is calculated for each of 15 samples from the same populations in Section 5.1. All the AKME curves are close to the true survival function.

are reported in the following two tables. Under the null hypothesis, that is $\theta = 1$, Table I lists the proportion of 1000 simulations with p-value $\leqslant p_0$ for the five testing methods: the weighted log-rank (WL) test proposed in Section 4, the bootstrap (BS) test, the test based on the proportional hazards model (PH), the standard log-rank test based on the Kaplan–Meier estimate (KM), and the stratified test (ST) [21] combining the standard log-rank test statistics over the strata. Except for the KM test, all tests perform well with both a low and a high amount of censoring. Table II shows the power for the four tests WL, BS, PH and ST under two alternative hypotheses $\theta = 1.25$ and 1.5. Because the survival times are generated by proportional hazards models, a high power of the PH test is expected. However, the tests WL, BS and ST do well, considering their non-parametric features.

5.3. Data with non-proportional hazards

The last example shows the performance of AKME and the hypothesis testing when the proportional hazards model is invalid. Suppose that the hazard rate functions of the treatment and control groups are the same defined by

$$h^1(t,z) = h^0(t,z) = 0.2 + z^2$$

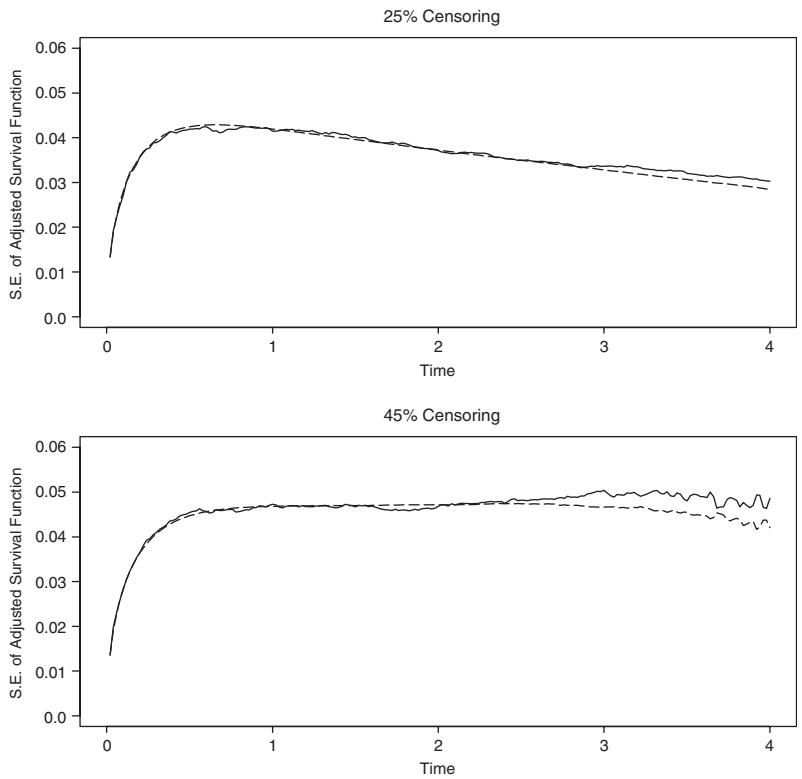

Figure 3. Standard deviation of the AKME calculated by Formula (5) in dashed lines and by Monte Carlo method in solid lines.

Table I. Performance of tests in terms of proportion of 1000 simulations with p-value $\leqslant p_0$ under H_0.

Under H_0	25 per cent censor					45 per cent censor				
p_0	WL	BS	PH	KM	ST	WL	BS	PH	KM	ST
0.010	0.009	0.011	0.011	0.908	0.011	0.012	0.009	0.012	0.857	0.009
0.025	0.026	0.025	0.024	0.950	0.026	0.031	0.024	0.030	0.913	0.029
0.050	0.053	0.048	0.061	0.974	0.060	0.057	0.051	0.057	0.956	0.056
0.100	0.104	0.107	0.102	0.983	0.103	0.109	0.106	0.106	0.981	0.107

WL = weighted log-rank test, BS = bootstrap test, PH = test based on proportional hazards model, KM = standard log-rank test based on Kaplan-Meier estimator, ST = stratified test.

where z is a covariate with value 0, 1 or -1. Assume the covariate follows the distribution

$$p(z) = \begin{cases} 1/2, & z = 0 \\ 1/4, & z = \pm 1 \end{cases}$$

Table II. Power of four tests in terms of proportion of 1000 simulations with p-value $\leqslant p_0$ under two alternative hypotheses.

	25 per cent censor				45 per cent censor			
p_0	WL	BS	PH	ST	WL	BS	PH	ST
$H_a : \theta = 1.25$								
0.010	0.052	0.058	0.091	0.098	0.053	0.054	0.063	0.065
0.025	0.111	0.118	0.168	0.169	0.097	0.107	0.116	0.111
0.050	0.194	0.195	0.250	0.250	0.162	0.177	0.195	0.189
0.100	0.285	0.280	0.343	0.340	0.253	0.258	0.287	0.272
$H_a : \theta = 1.50$								
0.010	0.254	0.243	0.329	0.332	0.198	0.201	0.262	0.264
0.025	0.366	0.358	0.447	0.451	0.307	0.303	0.373	0.372
0.050	0.481	0.473	0.562	0.555	0.405	0.398	0.479	0.477
0.100	0.617	0.595	0.692	0.687	0.506	0.500	0.602	0.598

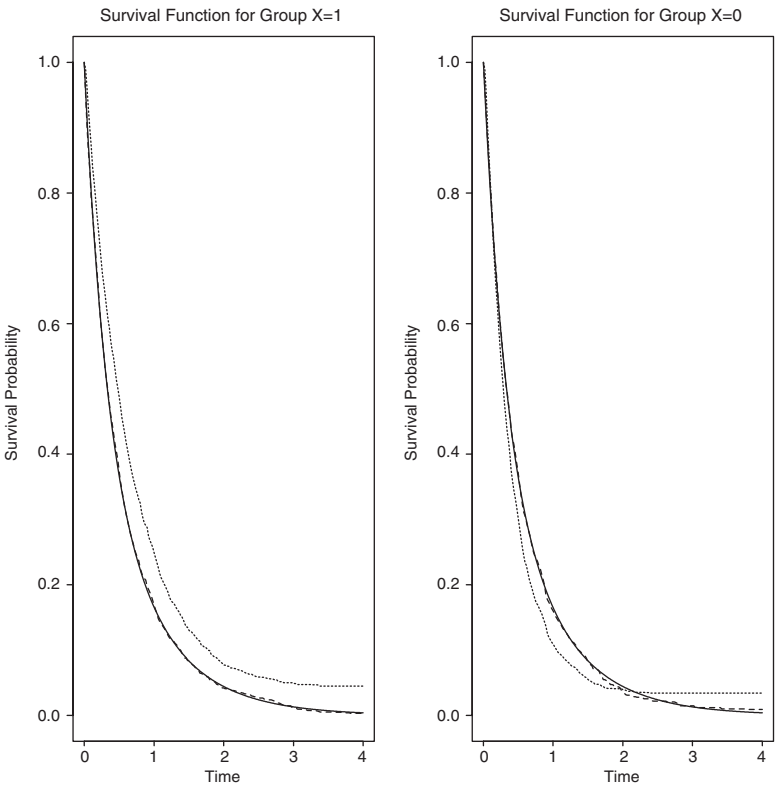

Figure 4. Comparisons of survival estimators by the AKME and the expected survival curves for the data with non-proportional hazards. The true survival curve (—-); the AKME (- - -); the expected survival curve (…).

Moreover, we have a non-randomized study design. The probability of being in the treatment group depends on z through the function

$$p(x=1|z)=\begin{cases}3/4, & z=0\\ 1/4, & z=\pm 1\end{cases}$$

The censoring time is from an exponential distribution with mean 3, which corresponds to a 15 per cent censoring rate.

A sample of size 200 is generated from the above model. The probability of being in the treatment group is estimated using the approach of maximum-likelihood estimate. Then the AKME of survival functions are obtained by Formula (1). We compare AKME with several other approaches by testing the null hypothesis that the treatment and control groups have the same survival rate. In addition to the WL test, the regular log-rank test based on the Kaplan–Meier estimator and the ST is conducted. A Cox model is also fitted to show the deviation of the model from the non-proportional data. The significance level is set at 0.05. In 1000 simulations of the hypothesis testing, each of which is based on a sample of size 200 from the previous model, the proportions of rejecting H_0 are, Cox model 82.2 per cent, Kaplan–Meier 82.4 per cent, ST 5.4 per cent, and the WL test of the AKME 5.1 per cent. In this example, the AKME and the stratification method provide comparable estimates of the treatment and control groups.

As an illustration, the AKME is also compared with the expected survival curve, which is the average of the survival estimates using the Cox proportional hazards model. Twenty-five samples of size 200 are generated from the same non-proportional hazards model. The average of the 25 AKMEs and the expected survival curves are shown in Figure 4, respectively. In this example where the proportional hazards assumption is invalid, the AKME provides a better estimator, closer to the true survival function.

6. APPLICATIONS

6.1. Times to reinfection for patients with sexually transmitted diseases

A study has been conducted to characterize times to reinfection for patients with sexually transmitted diseases (STD). A brief description of the STD data can be found in Klein and Moeschberger's book [21]. A sample of 877 individuals, with an initial diagnosis of one of the two STDs, gonorrhea and chlamydia, was followed for reinfection. In addition to the time to reinfection, 16 other variables were measured, which can be classified as demographic variables, behavioural factors, and symptom indicators. For example, the demographic variables include race, marital status, age of patient at initial infection, years of schooling, and type of initial infection. We want to study the race effect to determine whether it is an important factor to the reinfection times. In this data set, 33 per cent are white and 67 per cent are black. The survival functions of the two race groups within two years are estimated and compared.

The survival estimators using four different methods are shown in Figure 5: the regular Kaplan–Meier estimator, the AKME, a stratified Kaplan–Meier estimator as described below, and the expected survival curve. To calculate the AKME, the probability of being a white person conditioning on all other 15 variables is estimated by a logistic regression. This

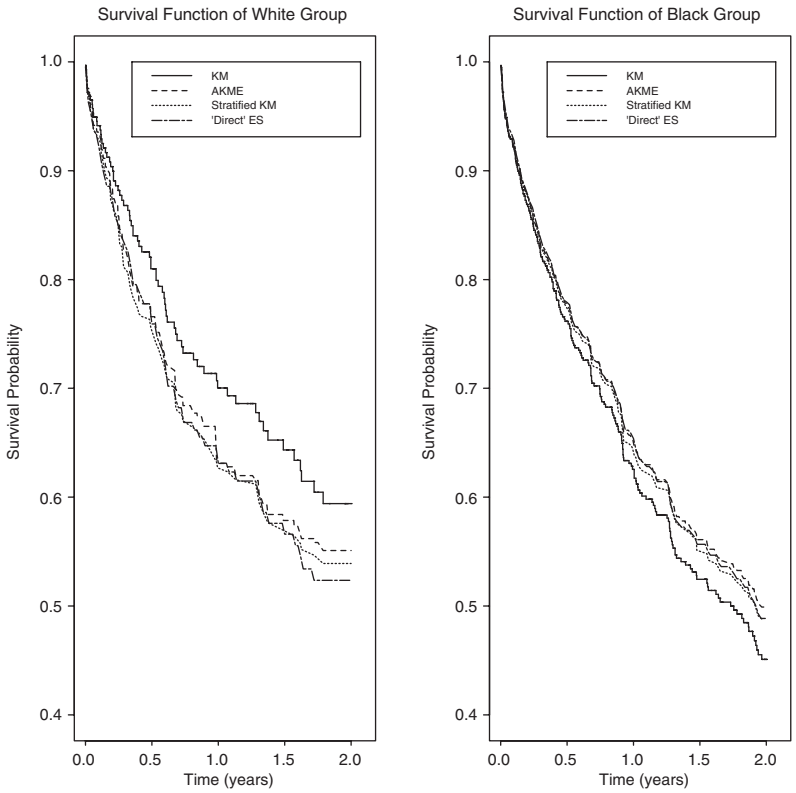

Figure 5. Comparisons of survival estimators by four different methods for the example of times to reinfection of the sexually transmitted diseases. The methods are Kaplan–Meier (KM); AKME; Stratified Kaplan–Meier (Stratified KM); expected survival curve ('Direct' ES). The survival curves of two race groups are estimated. Only the Kaplan–Meier estimators suggest a difference between the survival rates of the two race groups.

conditional probability, or propensity score according to Rosenbaum and Rubin's definition [16], is also used to obtain the stratified Kaplan–Meier estimator in Figure 5. Five strata are used. The subjects with similar propensity scores are matched together. The average of the five Kaplan–Meier estimators in each stratum gives the stratified Kaplan–Meier estimator.

The Kaplan–Meier curves suggest an improved survival for patients in the white group over those in the black group. However, when using the AKME, the expected survival curve, or the stratified Kaplan–Meier estimator, to adjust for other covariates, there is no longer a significant survival difference. The hypothesis tests for comparing the survival functions of the white and black groups confirm the results. The following p-values are obtained: the standard log-rank test 0.085; the WL test 0.599; the ST 0.464; and the test for zero coefficient of race based on the Cox model 0.745. Therefore, we conclude that race is not an important factor for the time to reinfection of STD.

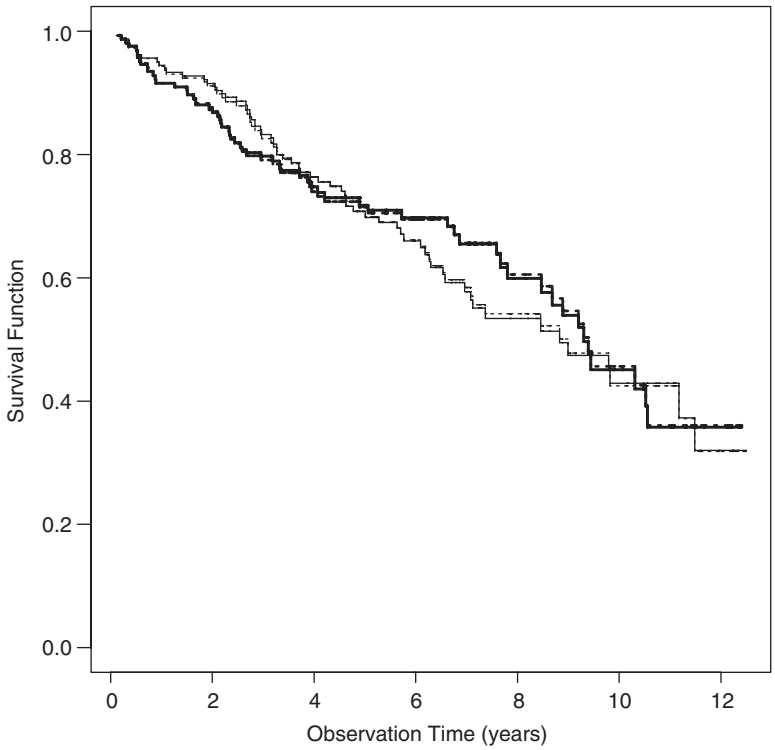

Figure 6. Estimated survival curves for the DPCA (light lines) and placebo (dark lines) groups by the AKME (solid lines) and the Kaplan–Meier (dashed lines).

6.2. Gender effect in the PBC data

The AKME and the WL test are applied to the data from primary biliary cirrhosis of the liver (PBC) conducted between 1974 and 1984 by the Mayo Clinic [22, 23]. PBC is a chronic liver disease of unknown cause and is fatal unless the patient receives a liver transplant. There are 424 patients in the Mayo Clinic trial and 312 of them participated in the randomized trial and were randomly assigned to either a placebo or the drug D-penicillamine (DPCA) group. Some previous statistical analyses of PBC data [23] suggested that the drug (D-penicillamine) is not effective in treatment of PBC disease, but bilirubin and age are important covariates which affect the hazard rate function.

In this analysis, we use the data from the 312 patients who participated in the randomized trial. Even though the trial was randomized, the analysis can be improved by a more refined approach that considers covariates. The survival functions of the DPCA and placebo groups are calculated using AKME. The probabilities of treatment assignment are estimated using logistic regression with six variables: age, sex, bilirubin, prothrombin time, albumin, and edema, with logarithm transformation for bilirubin, prothrombin time, and albumin. Figure 6 shows the survival curves of the AKME and the regular Kaplan–Meier estimators of the DPCA and placebo groups. Notice that the AKME curves are close to the regular Kaplan–Meier

estimators, which is an anticipated result because the data are randomized to the two groups. A group comparison test is conducted. The WL test gives a p-value of 0.684, suggesting that there is no significant difference in survival functions of DPCA and placebo groups.

In addition to the interests in the treatment effects, we can use the PBC data to study other features of the disease. For example, it is reported that PBC is more frequently a disease of women than men. Female patients are typically between 40 and 65 years of age but male patients are older. Men who are affected have a disease cause similar to that of women but appear to develop hepatocellular carcinoma later. Some researchers suspected that being male may pose a higher risk for severity than being female. We compare men and women to study the effect of sex on the survival rate of PBC.

The Kaplan–Meier estimates of the sex groups and the corresponding log-rank test suggest that male patients have a higher hazard risk with a significant p-value of 0.039. The Cox proportional hazards model with the seven variables drug, age, sex, bilirubin, prothrombin time, albumin, and edema, without logarithm transformation for bilirubin, prothrombin time, and albumin, supports the difference with a p-value of 0.019 for testing the parameter of sex. However, research showed that the logarithm transformations of bilirubin, prothrombin time, and albumin are necessary to satisfy the proportionality condition of the Cox model. After these transformations the Cox model gives a p-value of 0.149 for testing the parameter of sex. It suggests that the two gender groups cannot be separated.

Note that although the trial was randomized on DPCA and placebo groups, the sex groups were not randomized. The AKME of the survival functions for the two gender groups are obtained using five covariates to calculate the conditional probability of being in the male group: age, bilirubin, prothrombin time, albumin, and edema, without logarithm transformation for bilirubin, prothrombin time, and albumin. The WL test gives a p-value of 0.314. Furthermore, the AKME is recalculated using the five covariates but with logarithm transformation for bilirubin, prothrombin time, and albumin. The WL test gives a p-value of 0.287.

For more comparisons, we test for a sex effect with stratified estimators. With five continuous covariates, direct stratification is impracticable. Again, the observations are matched using the propensity score, the probability of being a male conditioning on the five covariates age, edema, and the logarithms of bilirubin, prothrombin time, and albumin. Five strata are used. The stratified test gives a p-value of 0.308, which supports the conclusion that there is no significant difference in survival functions between women and men groups.

In Figure 7, the Kaplan–Meier estimators of the female and male survival curves are separate, whereas the AKME of the two groups are closer. The AKME reduces the confounding effect of age and other covariates, and therefore provides a better estimation of survival functions for female and male groups. The WL test gives the same result with or without log transformation of the covariates. It also matches the result from a well-checked Cox model. The AKME and the WL test provide a more stable method for group comparison in this example.

7. DISCUSSION

Estimation and comparison of survival functions cannot be easily addressed by means of standard procedures when the number of covariates is large and their effects on the treatment and survival rate are not known precisely. The AKME and the weighted log-rank test proposed

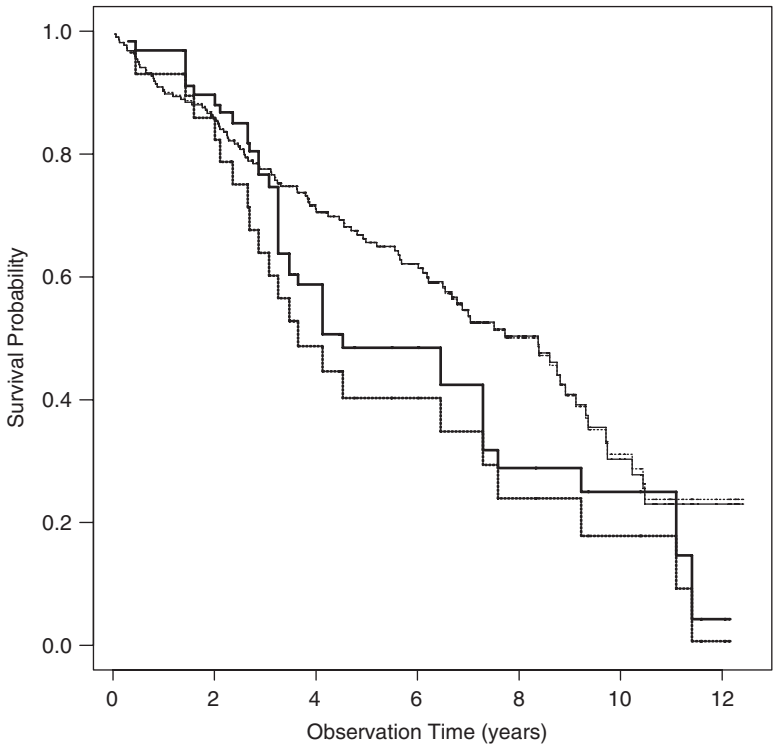

Figure 7. Estimated survival curves for the women (thin lines) and men (thick lines) groups by the AKME (solid lines) and the Kaplan–Meier (dashed lines) methods.

here adjust for confounding by using propensity scores to construct weights for individual observations. It is easy to calculate, provides marginal survival function estimates, and does not assume a parametric survival model. In contrast, the Cox proportional model requires model assumptions; matching or stratification has technical problems with a large set of confounding variables and the results of stratified estimation may vary for different subclassification strategies.

The definitions of treatment and control groups should be interpreted in a general way. The AKME and the weighted log-rank test can be used to estimate and compare any two groups. For example, as we did in the PBC example, the male and female groups were compared. In that case, the conditional probability of treatment is the conditional probability of a subject being male (or female) given the values of the covariates.

When the probability of a subject being in the treatment group is known, the AKME is a consistent estimate and its variance is provided by Formula (4). When the probability is unknown, it may be estimated by either fitting a parametric model of X_i over $\mathbf{Z}_i$, such as logistic regression, or by a non-parametric approach. The variance of the AKME can be estimated by Formula (5). It can be shown that if the probability estimators uniformly converge to the true probabilities, the AKME is still consistent. Moreover, Figure 3 displayed that Formula (5) provides a good variance estimate even though the estimated $\hat{p}_i$ is used.

APPENDIX A

A.1. Derivation of formula (2) for the pseudo-likelihood

For subject i assigned to the treatment group

$$
\begin{aligned}
& E\left\{\frac{X_i}{p_i}[\delta_i \log(S(T_i-0)-S(T_i))+(1-\delta_i)\log(S(T_i))]\right\} \\
&= E\left\{\left(E\left\{\frac{X_i}{p(\mathbf{Z}_i)}[\delta_i \log(S(T_i-0)-S(T_i))+(1-\delta_i)\log(S(T_i))]\right\}\middle|\mathbf{Z}=\mathbf{Z}_i\right)\right\} \\
&= E\left\{E\left([\delta_i \log(S^1(T_i-0)-S^1(T_i))+(1-\delta_i)\log(S^1(T_i))]|\mathbf{Z}=\mathbf{Z}_i\right)\right. \\
&\quad \left.\cdot\frac{E(X_i|\mathbf{Z}=\mathbf{Z}_i)}{p(\mathbf{Z}_i)}\right\} \\
&= E\{E([\delta_i \log(S^1(T_i-0)-S^1(T_i))+(1-\delta_i)\log(S^1(T_i))]|\mathbf{Z}=\mathbf{Z}_i)\} \\
&= E\{\delta_i \log(S^1(T_i-0)-S^1(T_i))+(1-\delta_i)\log(S^1(T_i))\}
\end{aligned}
$$

where the first equality uses double expectation and the second equality uses conditional independence of X_i and the survival processes, and the independence of X_i and δ_i. Summing over all subjects in the treatment group, we conclude the equality in Formula (2).

A.2. Proof of Proposition 1

Proof

To show the maximum pseudo-likelihood estimate of the survival function $S^1(t)$, we rewrite the pseudo-likelihood following the settings of Kaplan and Meier [17]. More specifically, let $t_1<t_2<\cdots<t_k$ denote the distinct times of death in the treatment group. Let λ_j denote the number of censored subjects in the interval $[t_j, t_{j+1})$, including the censors at t_j but not at t_{j+1}, and set $t_0=0$, $t_{k+1}=\infty$. The censoring time of the λ_j subjects are denoted by $L_i^{(j)}$, and their probabilities of being assigned to treatment are $p_i^{(j)}, i=1,2,\ldots,\lambda_j$. Then the pseudo-likelihood of the data can be written as

$$
\begin{aligned}
&\prod_{i=1}^{\lambda_0}[S^1(L_i^{(0)})]^{\frac{1}{p_i^{(0)}}} \\
&\quad\times[S^1(t_1-0)-S^1(t_1)]^{\sum_{i:T_i=t_1}\frac{\delta_i}{p_i}}\cdot\prod_{i=1}^{\lambda_1}[S^1(L_i^{(1)})]^{\frac{1}{p_i^{(1)}}} \\
&\quad\vdots \\
&\quad\cdot[S^1(t_k-0)-S^1(t_k)]^{\sum_{i:T_i=t_k}\frac{\delta_i}{p_i}}\cdot\prod_{i=1}^{\lambda_k}[S^1(L_i^{(k)})]^{\frac{1}{p_i^{(k)}}} \qquad \text{(A1)}
\end{aligned}
$$

where p_i and $p_i^{(j)}$ are positive probabilities of treatment.

It is easy to check that the pseudo-likelihood function (A1) is maximized by making $S^1(L_i^{(0)}) = 1$ and

$$S^1(t_j) = S^1(L_i^{(j)}) = S^1(t_{j+1} - 0)$$

for all i and j. We denote the common value above by S_j^1. Then the pseudo-likelihood (A1) becomes

$$\prod_{j=1}^{k} (S_{j-1}^1 - S_j^1)^{\sum_{i:T_i=t_j} \delta_i/p_i} \cdot (S_j^1)^{\sum_{i=1}^{\lambda_j} 1/p_i^{(j)}} \text{ with } S_0 = 1 \tag{A2}$$

Writing $s_j^1 = 1 - v_j^1 = S_j^1/S_{j-1}^1$, we get $S_j^1 = s_1^1 s_2^1 \cdots s_j^1$ and $S_{j-1}^1 - S_j^1 = s_1^1 \cdots s_{j-1}^1 v_j^1$. Substituting these to Formula (A2), the pseudo-likelihood further becomes

$$\prod_{j=1}^{k} (s_j^1)^{(\sum_{i:T_i \geqslant t_j} 1/p_i) - (\sum_{i:T_i=t_j} \delta_i/p_i)} \cdot (1 - s_j^1)^{\sum_{i:T_i=t_j} \delta_i/p_i} \tag{A3}$$

where we utilize the fact that

$$\sum_{i:T_i \geqslant t_j} \frac{1}{p_i} = \sum_{l=j}^{k} \left(\sum_{i=1}^{\lambda_l} \frac{1}{p_i^{(l)}} + \sum_{i:T_i=t_l} \frac{\delta_i}{p_i} \right) \quad \text{for } j = 1, \ldots, k$$

In Formula (A3) each s_j^1 is maximized individually by the binomial estimate

$$\hat{s}_j^1 = \frac{\sum_{i:T_i \geqslant t_j} 1/p_i - \sum_{i:T_i=t_j} \delta_i/p_i}{\sum_{i:T_i \geqslant t_j} 1/p_i} = 1 - \frac{\sum_{i:T_i=t_j} \delta_i/p_i}{\sum_{i:T_i \geqslant t_j} 1/p_i} = 1 - \frac{d_{j1}^w}{Y_{j1}^w}$$

Therefore the survival function is estimated by

$$\hat{S}^1(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_j \leqslant t} \left[1 - \dfrac{\sum_{i:T_i=t_j} \delta_i/p_i}{\sum_{i:T_i \geqslant t_j} 1/p_i} \right] & \text{if } t_1 \leqslant t \end{cases}$$

It is the AKME defined by Formula (1). □

A.3. Proof of Proposition 2

Proof

(a) We first derive the variance estimation when the probabilities of treatment p_i's are assumed known. Consider the conditional expectation and variance for $\hat{s}_j^1$ given information up to time t_j. Then

$$E[\hat{S}^1(t)]^2 = \prod_{j=1}^{l} E_j(\hat{s}_j^1)^2 = \prod_{j=1}^{l} ((s_j^1)^2 + \text{var}_j(\hat{s}_j^1))$$

$$= \prod_{j=1}^{l} \left((s_j^1)^2 + \frac{s_j^1(1-s_j^1)}{M_j} \right) = (s_1^1)^2 \cdots (s_l^1)^2 \prod_{j=1}^{l} \left(1 + \frac{1-s_j^1}{M_j s_j^1} \right)$$

$$= (S^1(t))^2 \prod_{j=1}^{l} \left(1 + \frac{1-s_j^1}{M_j s_j^1} \right)$$

where

$$\frac{1}{M_j} = \frac{\sum_{i:T_i \geqslant t_j} (1/p_i)^2}{(\sum_{i:T_i \geqslant t_j} 1/p_i)^2}$$

Under the condition that

$$\frac{\max_{i:T_i \geqslant t_j} (1/p_i)}{\sum_{i:T_i \geqslant t_j} 1/p_i} \to 0$$

we would have $M_j^{-1} \to 0$ and therefore may ignore terms of order M_j^{-2}. The variance is approximated by

$$\mathrm{Var}[\hat{S}^1(t)] = (S^1(t))^2 \left[\prod_{j=1}^{l} \left(1 + \frac{1-s_j^1}{M_j s_j^1} \right) - 1 \right] \approx (S^1(t))^2 \sum_{j=1}^{l} \frac{1-s_j^1}{M_j s_j^1}$$

(b) Denote $\hat{\mathbf{p}} = \{\hat{p}_i, i = 1, \ldots, n\}$ as the estimated probabilities of treatment given the covariate vector $\mathbf{Z}_i$, $i = 1, \ldots, n$. The estimate $\hat{\mathbf{p}}$ is a random vector determined by $(X_i, \mathbf{Z}_i)$, $i = 1, \ldots, n$. Now the variance of the AKME can be written as

$$\mathrm{Var}[\hat{S}^1(t)] = E(\mathrm{Var}[\hat{S}^1(t)|\hat{\mathbf{p}}]) + \mathrm{Var}(E[\hat{S}^1(t)|\hat{\mathbf{p}}])$$

In the first term, $\mathrm{Var}[\hat{S}^1(t)|\hat{\mathbf{p}}]$ is obtained from (a) when p_i's are given. In the second term, $E[\hat{S}^1(t)|\hat{\mathbf{p}}] = S^1(t)$ for $t < T_{\max}$, which is a non-random quantity, hence $\mathrm{Var}(E[\hat{S}^1(t)|\hat{\mathbf{p}}]) = 0$. In fact, as implied in the derivation of the mean of AKME in Formula (3), $E_j(1-\hat{s}_j^1)$ does not depend on the values of p_i's. Therefore, $E(1-\hat{s}_j^1|\hat{\mathbf{p}}) = 1-s_j^1$ and consequently $E[\hat{S}^1(t)|\hat{\mathbf{p}}] = S^1(t)$ for $t < T_{\max}$. The variance

$$\mathrm{Var}[\hat{S}^1(t)] = E(\mathrm{Var}[\hat{S}^1(t)|\hat{\mathbf{p}}]) = E\left((S^1(t))^2 \sum_{j:t_j \leqslant t} \frac{1-s_j^1}{\hat{M}_j s_j^1} \right)$$

where $\hat{M}_j$ is calculated at $\hat{p}_i$, $i = 1, \ldots, n$, and the expectation is in term of the estimated probabilities $\hat{\mathbf{p}}$. A reasonable estimation of this expectation is

$$[\hat{S}^1(t)]^2 \sum_{j:t_j \leqslant t} \frac{1-\hat{s}_j^1}{\hat{M}_j \hat{s}_j^1}$$

which is Formula (5). □

A.4. Variance derivation of the log-rank statistic

To calculate the variance of the statistic G^w, we apply a result of moment computations for the sum of a random sample from zero-sum scores [24]. In fact, at time t_j the term $d_{j1}^w - Y_{j1}^w(d_j^w/Y_j^w)$ can be written as the sum of a random sample of size d_j chosen without replacement from the Y_j zero-sum scores $U_1,\ldots,U_{Y_j}$. The zero-sum scores are defined as

$$U_i = \begin{cases} w_i - \dfrac{Y_{j1}^w}{Y_j^w} w_i & \text{if } X_i = 1 \\ -\dfrac{Y_{j1}^w}{Y_j^w} w_i & \text{if } X_i = 0 \end{cases}$$

for $i = 1,\ldots,Y_j$. It is easy to check that

$$\begin{aligned} \sum_{i=1}^{Y_j} U_i &= \sum_{i=1}^{Y_j} \left(w_i - \frac{Y_{j1}^w}{Y_j^w} w_i \right) X_i + \sum_{i=1}^{Y_j} \left(-\frac{Y_{j1}^w}{Y_j^w} w_i \right) (1 - X_i) \\ &= \sum_{i=1}^{Y_j} w_i X_i + \sum_{i=1}^{Y_j} \left(-\frac{Y_{j1}^w}{Y_j^w} w_i \right) \\ &= Y_{j1}^w - \frac{Y_{j1}^w}{Y_j^w} \cdot Y_j^w = 0 \end{aligned}$$

We also have

$$\begin{aligned} V &= \sum_{i=1}^{Y_j} U_i \cdot \delta_i I(T_i = t_j) = \sum_{i:T_i=t_j} w_i X_i \delta_i - \sum_{i:T_i=t_j} \frac{Y_{j1}^w}{Y_j^w} w_i \delta_i \\ &= d_{j1}^w - Y_{j1}^w \left(\frac{d_j^w}{Y_j^w} \right) \end{aligned}$$

Therefore, $d_{j1}^w - Y_{j1}^w(d_j^w/Y_j^w)$ is the sum of a sample of size $d_j = \sum_{i=1}^{Y_j} \delta_i I(T_i = t_j)$ chosen without replacement from the zero-sum scores $U_1,\ldots,U_{Y_j}$. According to Peto and Peto [24], the conditional moments of $d_{j1}^w - Y_{j1}^w(d_j^w/Y_j^w)$ given the information up to time t_j can be calculated by formulas of $E(V) = 0$ and

$$\begin{aligned} \mathrm{Var}(V) &= \frac{d_j(Y_j - d_j)}{(Y_j - 1)} \sum_{i=1}^{Y_j} \frac{U_i^2}{Y_j} \\ &= \frac{d_j(Y_j - d_j)}{Y_j(Y_j - 1)} \sum_{i=1}^{Y_j} \left[\left(\frac{Y_{j0}^w}{Y_j^w} \right)^2 w_i^2 X_i + \left(\frac{Y_{j1}^w}{Y_j^w} \right)^2 w_i^2 (1 - X_i) \right] \end{aligned}$$

Using double expectation and the conditional variance formula above, we obtain the variance of statistic G^w.

ACKNOWLEDGEMENTS

The authors thank the editor and the anonymous referees for their helpful comments and suggestions that help enhance the results.

REFERENCES

1. Cox DR. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society B* 1972; **24**:187–220.
2. Aalen OO. A linear regression for the analysis of life times. *Statistics in Medicine* 1989; **8**:907–925.
3. Thomsen BL, Keiding N. A note on the calculation of expected survival, illustrated by the survival of liver transplant patients. *Statistics in Medicine* 1991; **10**:733–738.
4. Therneau TM, Grambsch PM. *Modeling Survival Data*: *Extending the Cox Model*. Springer: New York, 2000.
5. Hankey BF, Myers MH. Evaluating differences in survival between two groups of patients. *Journal of the Chronic Diseases* 1971; **24**:523–531.
6. Cupples LA, Gagnon DR, Ramaswamy R, D'Agostino RB. Age-adjusted survival curves with application in the Framingham study. *Statistics in Medicine* 1995; **14**:1731–1744.
7. Amato DA. A generalized Kaplan–Meier estimator for heterogeneous populations. *Communication in Statistics-Theory and Methods* 1988; **17**:263–286.
8. Nieto FJ, Coresh J. Adjusting survival curves for confounders: a review and a new method. *American Journal of Epidemiology* 1996; **143**:1068–1069.
9. Winnett A, Sasieni P. Adjusted Nelson–Aalen estimates with retrospective matching. *Journal of the American Statistical Association* 2002; **97**:245–256.
10. Galimberti S, Sasieni P, Valsecchi MG. A weighted Kaplan–Meier estimator for matched data with application to the comparison of chemotherapy and bone-marrow transplant in leukaemia. *Statistics in Medicine* 2002; **21**:3847–3864.
11. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 1951; **47**:663–685.
12. Zhao LP, Lipsitz S. Design and analysis of two-stage studies. *Statistics in Medicine* 1992; **11**:769–782.
13. Wang CY, Wang S, Zhao LP, Ou ST. Weighted semiparametric estimation in regression analysis with missing covariate data. *Journal of the American Statistical Association* 1997; **92**:512–525.
14. Dawson R, Lavori RW. Using inverse weighting and predictive inference to estimate the effects of time-varying treatments on the discrete-time hazard. *Statistics in Medicine* 2002; **21**:1641–1661.
15. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine* 2004; **23**:2937–2960.
16. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**:41–55.
17. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958; **53**:457–481.
18. Kalbfleisch JD, Lawless JF. Likelihood analysis of multi-state models for disease incidence and mortality. *Statistics in Medicine* 1988; **7**:149–160.
19. Flanders WD, Greenland S. Analytic methods for two-stage case-control studies and other stratified designs. *Statistics in Medicine* 1991; **10**:739–747.
20. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer: New York, 1993; 257–259.
21. Klein JP, Moeschberger ML. *Survival Analysis*: *Techniques for Censored and Truncated Data* (2nd edn). Springer: New York, 2003.
22. Dickson ER, Grambsch PM, Fleming TR, Fisher LD, Langworthy A. Prognosis in primary biliary cirrhosis: model for decision making. *Hepatology* 1989; **10**:1–7.
23. Fleming TR, Harrington DP. *Counting Processes and Survival Analysis*. Wiley: New York, 1991.
24. Peto R, Peto J. Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society A* 1972; **135**:185–207.

