Structural Models for Analyzing Survival Data with Multiple Time-dependent Treatments

Yeqian Liu[*], Yingxiao Huang

Department of Mathematical Sciences, Middle Tennessee State University, Murfreesboro, USA

Abstract   In randomized clinical trials involving survival time, patients may initiate secondary salvage treatments during the follow-up in order to prolong survival. As a result, the effect of primary treatment is usually confounded by these secondary treatments. In addition, such secondary salvage treatments are usually time-dependent due to being initiated upon adverse reaction, disease progression, etc. In this paper, we adopt two types of structural models: structural nested models and marginal structural Cox models, and propose an inferential procedure that improves the efficiency of the usual IPTW method. We examine the finite-sample performance of the proposed method by simulation studies and by application to data from a cancer clinical trial.

Keywords   Marginal structural model, Inverse probability weighting, Causal inference, Bivariate log-rank test, Artificial Censoring

1. Introduction

In randomized clinical trials, an improvement in patients' survival time is considered the most convincing measure of drug efficacy and clinical benefit. This article is motivated by a randomized phase III clinical trial comparing Cisplatin (control) and Vinorelbine/Cisplatin combination (experimental) in patients with non-small tumor lung cancer in terms of the time to all-cause death. It was found that many patients initiated a secondary salvage chemotherapy during the follow-up after discontinuing from the study treatment. More patients in the experimental treatment arm received a secondary treatment, and the time from discontinuation of the study treatment to initiation of a secondary treatment was considerably shorter. The main goal of applying secondary treatments is to prolong patients' survival time. However, the imbalance of secondary treatment between control and experimental arms might biased the estimation of primary treatment effect. Therefore, it is important to estimate the effect of the primary treatment under the condition of balanced secondary treatment between the two groups, namely, the direct effect [1].

The initiation time of a secondary treatment is not stated in the protocol of the clinical trial but rather depend on the choices of the patients or physicians. Such decisions usually depend on patients' baseline or post-randomization disease condition, which could potentially confound the effect of the primary treatment. Common analysis tools such as landmark analysis, time-dependent Cox models, and censoring subjects at the beginning of secondary treatment do not perform well on estimating primary treatment effect. Two types of structural models: the marginal structural Cox model [2,3] and the structural nested accelerated failure time model with the inferential method of g-estimation [5] have been developed for estimating effect of time-dependent treatments in observational studies. Both causal modes are built upon the counterfactual framework. In this paper, we adopt these two models for making causal inference, and propose an efficient inferential method in the setting of a randomized trial with a secondary treatment.

The remainder of the article is organized as follows. Section 2 set ups the notation and describes the proposed method and related background. Simulation studies and analysis of the motivating study are reported in Sections 3 and 4, respectively. The article concludes with a discussion on the choice of models under different scenarios in Section 5.

2. Models and Inference Procedure

2.1. Structural Nested Failure Time Models

2.1.1. Log-rank Test

Log-rank test is one of the most commonly used method for comparing survival functions between two treatment groups. Let $j = 1, \dots, J$ be the distinct observed failure

times. For each time j, let N_{0j} and N_{1j} be the number of subjects at risk at the start of period j in the treatment and control arm, respectively. Let $N_j = N_{0j} + N_{1j}$. Let O_{0j} and O_{1j} be the observed number of events in the arms respectively, and define $O_j = O_{0j} + O_{1j}$. Conditional on O_j, the quantity O_{1j} has a hypergeometric distribution with parameters N_j, N_{1j} and O_j, whose expectation

$$E_{1j} = \frac{O_j}{N_j} N_{1j}$$

and the variance

$$V_j = \frac{O_j(N_{1j}/N_j)(1 - N_{1j}/N_j)(N_j - O_j)}{N_j - 1}.$$

The regular log rank test statistic is given by

$$Q_L = \frac{\{\sum_j (O_{1j} - E_{1j})\}^2}{\sum_j V_j},$$

which follows a χ_1^2 distribution under null hypothesis.

Extensions of the regular log rank test have been developed to increase power under specific circumstances. The extension is done via weighing each period j by a factor w_j. Specifically, the weighted log rank test statistic is given by

$$Q_W = \frac{\{\sum_j w_j (O_{1j} - E_{1j})\}^2}{\sum_j w_j^2 V_j},$$

in which the weight function w_j is assumed to be non-increasing. Below is a table that lists some commonly used weight functions.

Table 1. Commonly Used Weighted Log-rank Tests

Test	Weight w_j
Logrank	$w_j = 1$
Gehan's Wilcoxon	$w_j = r_j$
Peto/Prentice	$w_j = \hat{S}(t_j)$
Fleming-Harrington	$w_j = [\hat{S}(t_j)]^\rho \ (\rho \geq 0)$
Tarone-Ware	$w_j = \sqrt{r_j}$

Under null hypothesis, Q_W also follows a χ_1^2 distribution. The log-rank test can be considered as a special case of the weighted log rank test by setting $w_j = 1$.

2.1.2. Rank-preserving Structural Nested Failure Time (RPSFT) Models

The RPSFT model relates T_i to a potential outcome $T_{i(0)}$ that would have been observed without treatment through a treatment effect [6]. Consider a simple rank-preserving structural nested failure time (RPSFT) model [6,8], which links potentially counterfactual $T_{i(0)}$ with observable Z_i, T_i and P_i,

$$T_{i(0)}(\theta_1, \theta_2) = exp(\theta_1 Z_i) [P_i + exp(\theta_2) (T_i - P_i)],$$

Where $T_{i(0)}$: is the counterfactual survival time of patient i in control arm and never received secondary treatment, Z_i is the randomized primary treatment which equals 1 if in treatment arm and P_i is the moment when secondary treatment is initiated if ever given, time of death otherwise.

The distribution of counterfactual survival time should be independent of Z_i. As a result, there should be no significant difference across the two primary treatment arms based on log-rank test. Following model described in [6,8], the log-rank test statistic would be bivariate

$$Q_{LW}(\theta_1, \theta_2)$$
$$= (S_L(\theta_1, \theta_2), S_W(\theta_1, \theta_2)) \Sigma_{LW}(\theta_1, \theta_2)^- (S_L(\theta_1, \theta_2), S_W(\theta_1, \theta_2))'$$

where S_L denotes log-rank score, S_W is the Prentice-Wilcoxon log-rank score, and Σ_{LW} is the joint estimated covariance matrix applied to the data $\{Z_i, T_i\}$. Under the null hypothesis, $Q_{LW}(\theta_1, \theta_2)$ follows a χ_2^2 distribution. Unlike the case in treatment crossover [4] in which a regular log rank test statistic Q_L was used, it is important to employ two distinct log rank tests. Using only one log rank test can introduce identifiability issue since there are two parameters θ_1, θ_2 to be estimated.

The estimated coefficients $(\hat{\theta}_1, \hat{\theta}_2)$ can be obtained by solving for (θ_1, θ_2) that minimizes bivariate log rank test statistic $Q_{LW}(\theta_1, \theta_2)$. In term of finding the minimizer, the naïve grid search root finding method is usually inefficient and computational intensive. In order to increase computational efficiency, here we employ the Frank-Wolfe non-linear optimization method to find estimates. Note that the bivariate log-rank statistics is a complex function, which does not have closed form gradient, and the Frank-Wolfe optimization method does not require gradient.

2.2. Marginal Structure Cox Models

Marginal structural Cox models [2] was first proposed to draw valid causal inference from observational studies, in which the assignment of treatment and control/placebo is nonrandom and time-dependent treatment in the presence of confounders. The idea of marginal structure model is to inversely weigh each observation by the probability of receiving the treatment. Through the inverse probability of treatment weighting (IPTW), we create a pseudo-population in which the distributions of covariates are similar between two treatment groups and we can draw causal inference from the pseudo-population.

We consider the marginal structural Cox model

$$\lambda(t|Z) = \lambda_0(t) exp [\theta_1 Z + \theta_2 I(t)]$$

where Z represents the treatment and $I(t) = 1$ if secondary treatment is initiated at time t, $I(t) = 0$ otherwise. In the model, each observation is weighted by $\{P_i(t) \cdot P_i^*(t)\}$, where $P_i(t)$ corresponds to probability of receiving secondary treatment at time t, and $P_i^*(t)$ corresponds to uncensored probability at time t. Patients who have a extremely small probability of receiving secondary treatment will contribute a huge number of copies in the pseudo-population when taking the inverse of

the weight. Therefore, parameter estimates may become highly unstable and have large variability. To overcome this issue, we use stabilized weights $SP_i(t)$ and $SP_i^*(t)$, respectively.

In practice, the weights $P_i(t)$ can be estimated by fitting a Cox model that models the initiation of secondary treatment as an outcome. In the Cox model, we include both baseline and time-dependent covariates that potentially affects the initiation of secondary treatment. For example, when we estimate $P_i(t)$ in the treatment group where $Z = 1$, we fit the model

$$\lambda_i[t|Z_i = 1, X_i(t)] = \lambda_{01}(t)\, exp\big(\alpha_1' X_i(t)\big),$$

where $X_i(t)$ is a set of baseline and time-dependent covariates that are related to the initiation of secondary treatment and α_1 is the coefficient vector for $X_i(t)$. If patient i did not start secondary treatment at time t, then the estimated probability

$$\hat{K}_i(t) = \prod_{\{j:S_j<t, \Delta_j=1, Z_j=1\}} exp\big(-\hat{\lambda}_{01}(S_j)\, exp\{\hat{\alpha}_1' X_i(S_j)\}\big),$$

where $\hat{\lambda}_{01}(S_j) = \Delta_j / \sum_{i\in R} \exp\{\hat{\alpha}_1' X_i(S_j)\} Y_i(S_j)$ is the Breslow estimator of the baseline hazard in treatment arm and $Y_i(t)$ is the at-risk indicator. On the other hand, if the patient started secondary treatment, then the estimated probability at time $t \geq S_i$ is

$$\hat{K}_i(t) = \big[1 - exp\{-\hat{\lambda}_{01}(S_i)\, exp\{\hat{\alpha}_1' X_i(S_i)\}\}\big]$$
$$\times \prod_{\{j:S_j<S_i, \Delta_j=1, Z_j=1\}} exp\big(-\hat{\lambda}_{01}(S_j)\, exp\{\hat{\alpha}_1' X_i(S_j)\}\big),$$

and we have estimated weight $\hat{P}_i(t) = 1/\hat{K}_i(t)$. The stabilized weights can be obtained by repeating the above procedure with the covariates $X_i(t)$ removed from the Cox model and get estimated probabilities $\widehat{SK}_i(t)$. Then the stabilized weight $\widehat{SP}_i(t) = \widehat{SK}_i(t)/\hat{K}_i(t)$. We can also estimate the weights corresponding to censoring, $P_i^*(t)$ and $SP_i^*(t)$ by the similar procedure.

One important assumption for marginal structural model is all possible confounders must be included when estimating the weights. Obviously, there is no way to test this assumption. Therefore, it is recommended to include as many covariates in $\bar{L}(t)$ as possible in practice.

3. A Simulation Study

3.1. Data Generation

Now we present some results obtained from an extensive simulation study conducted to assess the performance of the estimation procedures proposed in the previous sections. For simplicity, we assume that the survival time and the initiation time of the secondary treatment follow exponential distributions. In specific, denote Z as the randomized primary treatment indicator, we generate the initiation time of the secondary treatment from the exponential distribution

$$T_2 = -\lambda_{0,2}^{-1} exp(-\beta Z - \gamma_1 X_1 - \gamma_2 X_2)\, log(u_2),$$

where $u_2 \sim Unif(0,1)$ and $\beta < 0$ so that primary treatment is beneficial on PFS. X_1 followed a Bernoulli distribution with probability 0.5, and X_2 was a standard normal random variable. We then generated the counterfactual survival time if the subject never received secondary treatment

$$T_1^{ZN} = -\lambda_{0,1}^{-1} exp(-\theta_1 Z)\, log(u_1).$$

We then generate X_2 from a normal distribution that is related to both T_2 and T_1, and hence X_2 is a confounder. A time-dependent variable $X_3(t)$ was also generated based on a Gaussian process.

Suppose that the secondary treatment was initiated at time V, we generated the observable survival time in the presence of secondary treatment via

$$T_{OS}^{ZT_2} = P + exp(-\theta_2)\, [T_{OS}^{ZV} - V].$$

Note that $\theta_2 < 0$ which indicated a beneficial secondary treatment. It can be seen that secondary treatment prolonged the remaining life from T_2 by a factor of $exp(-\theta_2)$. Note the expression for $T_{OS}^{ZT_2}$ holds only when T_{OS}^{ZN} has an exponential distribution.

We assume the censoring time for survival time and secondary treatment initiation time follow two independent uniform distributions. The patients entered the study at the rate of 34 per month, for 24 month. The total sample size was 816. The final analysis is conducted when we observed 451 deaths.

3.2. Models and Results

We compared the estimation of the treatment effects from the following four different models in our simulation study:

(1) Cox proportional hazards model with only primary treatment, which ignores secondary treatment.

(2) Cox proportional hazards model with both primary treatment and time-dependent secondary treatment, but ignores a confounder X_2 which affects both initiation of secondary treatment and survival time. In specific,

$$\lambda(t) = \lambda_0(t) exp\,(\theta_1 Z + \theta_2 I(t))$$

where I(t) equals 1 if subject is on secondary treatment at time t, 0 otherwise.

(3) RPSFT model. We set the starting value for the Frank-Wolfe optimization process to be $(\theta_1^{(0)} = \theta_{01}, \theta_2^{(0)} = 0)$, where θ_{01} is the estimated by the Cox model with time-dependent secondary treatment.

(4) Marginal structural model whose form is the same with time-dependent Cox model (2). Using inverse probability of treatment weighting (IPTW), each observation is weighted by the inverse probability estimated by fitting Cox models with covariates X_1, X_2 and $X_3(t)$. We also computed stabilized weights to control for extreme weights. Note that in the data generating step, X_2 is a confounder that relates to both survival time and initiation time of the secondary treatment. Therefore, the model was misspecified and by doing so we are also able to assess the robustness of the marginal structural model towards model misspecification.

Tables 2 and 3 gives the summarized simulation results. Table 3 considered the situation when primary treatment was

efficacious, table 2 studied the situation when there was no primary treatment effect. One can see from the table that when secondary treatment was highly effective compared to the primary treatment (table 2), ignoring the secondary treatment can introduce large amount of bias. In addition, Cox model with time-dependent secondary treatment showed accurate estimate of primary treatment, but the estimate of secondary treatment was very inaccurate. The RPSFT model also yielded accurate primary treatment estimate, and failed to estimate secondary treatment correctly. Furthermore, marginal structural model yielded was able to estimate both primary and secondary treatment quite accurately; and it was robust to model misspecification.

Table 2. Effective Primary Treatment

Model	Mean	Absolute Bias	Relative Bias	ESE	SD	95% CI coverage
Primary treatment effect log(HR) = -0.154 (HR = 0.85)						
Cox model	-0.479	0.325	2.110	0.151	0.148	0.391
Time-dependent Cox model	-0.165	0.011	0.071	0.135	0.126	0.964
RPSFT model	-0.174	0.020	0.130	0.165	0.141	0.972
Marginal Structural Model	-0.145	0.009	0.057	0.095	0.091	0.943
Marginal Structural Model (Stabilized Weight)	-0.148	0.006	0.041	0.092	0.088	0.951
Secondary treatment effect log(HR) = -1.204 (HR = 0.30)						
Cox model	N/A	N/A	N/A	N/A	N/A	N/A
Time-dependent Cox model	21.857	23.061	19.154	385.215	0.167	1.000
RPSFT model	1.052	2.256	1.874	54.613	0.581	1.000
Marginal Structural Model	-1.218	0.014	0.012	0.140	0.143	0.958
Marginal Structural Model (Stabilized Weight)	-1.212	0.012	0.010	0.136	0.135	0.953

Table 3. Ineffective Primary Treatment

Model	Mean	Absolute Bias	Relative Bias	ESE	SD	95% CI coverage
Primary treatment effect log(HR) = 0 (HR = 1)						
Cox model	-0.278	0.278	N/A	0.139	0.140	0.343
Time-dependent Cox model	-0.009	0.009	N/A	0.114	0.118	0.957
RPSFT model	-0.074	0.074	N/A	0.181	0.177	0.853
Marginal Structural Model	-0.005	0.005	N/A	0.102	0.105	0.952
Marginal Structural Model (Stabilized Weight)	-0.008	0.008	N/A	0.107	0.111	0.947
Secondary treatment effect log(HR) = -1.204 (HR = 0.30)						
Cox model	N/A	N/A	N/A	N/A	N/A	N/A
Time-dependent Cox model	23.533	24.738	20.547	354.713	0.161	1.000
RPSFT model	1.943	3.147	2.614	106.145	0.784	1.000
Marginal Structural Model	-1.223	0.019	0.016	0.136	0.132	0.952
Marginal Structural Model (Stabilized Weight)	-1.215	0.011	0.009	0.134	0.137	0.950

4. An Application

Now we apply the methods proposed in the previous sections to clinical trial for non-small tumor lung cancer. It was a randomized, double-blinded, multicenter phase III study in patients with lung cancer receiving Cisplatin (control) vs Vinorelbine/Cisplatin combination (experimental). The median survival time and initiation time of the secondary treatment in months are summarized in table 4.

In the marginal structural Cox model, we included six baseline covariates that are potentially linked to the initiation time of the secondary treatment. Further, we examined the distributions of those covariates (table 5). We can see that randomization is doing a good job in balancing the covariate distributions in treatment and placebo groups. The distribution of covariates were balanced in both control and experimental groups except for sex, where there were more

males in the experimental group.

Table 4.   Median Survival and Secondary Treatment Initiation Time

	Experimental	Control
Survival Time	23.6	28.2
Initiation Time of Secondary Treatment	15.0	9.2

The survival data was analyzed by five models: (1) Cox model ignoring secondary treatment; (2) time-dependent Cox model; (3) RPSFT model with Frank-Wolfe optimization root-finding; (4) Marginal structural model with non-stabilized weight; (5) Marginal structural model with stabilized weight. Results are summarized in table 5. In addition, we computed descriptive statistics of stabilized and non-stabilized weights in MSM (table 6). We can see that non-stabilized weights have a much larger maximum (613.581) compared with the maximum (1.684) of stabilized weights. Therefore, the stabilizing process was able to eliminate extreme large weights.

The primary treatment was shown to be ineffective on prolonging survival time by all five models. Only marginal structural model with stabilized weights detected a significant primary treatment effect. As for the secondary treatment effect, time-dependent Cox model and RPSFT models gave unreliable estimates, and the confidence intervals were too wide. A similar phenomenon was also observed by other studies [7]. One possible explanation of RPSFT models' unreliability to estimate secondary treatment effect can be seen via the surface plot (figure 1): there seemed to be a flat valley of θ_2 values along the section $-0.08 \leq \theta_1 \leq 0.04$ where the bivariate log-rank test statistic had uniformly small values. Therefore, the global minimum of the log-rank test statistic is hard to locate. On the other hand, the two marginal structural models yielded more reasonable estimate, indicating that the secondary treatment was highly effective.

Table 5.   Distribution of Patients' Baseline Covariates

	Experimental	Control	P Value[†]
Sex			
Male	135	97	0.0032
Female	85	116	
Primary Tumor Site			
Lung	125	131	0.6254
Other	96	82	
Race			
Caucasian	186	174	0.4185
Other	34	39	
Histologic Score			
Well-differentiated	136	145	
Moderately-differentiated	67	58	0.3411
Poorly-differentiated	3	2	
Unknown	14	8	
Non-target lesions?			
No	71	65	0.8716
Yes	149	148	
Age			
< 65	125	128	0.7154
≥ 65	95	85	
†: P value from χ^2 test comparing treatment and placebo groups			

Table 6.   Weights in Marginal Structural Model

Weight Type	Mean	Minimum	Q1	Median	Q3	Maximum
Non-stabilized	9.445	1	1.213	1.413	1.865	613.581
Stabilized	0.750	0.391	0.561	0.873	1.248	1.684

Table 7.   Estimates and Confidence Intervals

| Primary treatment effect				
Model	Log HR	HR Estimate	95% CI of HR	P Value
Cox model	0.091	1.095	(0.923, 1.257)	0.4651
Time-dependent Cox model	0.131	1.141	(0.973, 1.314)	0.1767
RPSFT model	0.081	1.205	(0.718, 1.684)	0.5619
Marginal Structural Model	0.052	1.128	(0.957, 1.281)	0.2355
Marginal Structural Model (Stabilized Weight)	0.070	1.074	(0.991 1.142)	0.0838
Secondary treatment effect				
Model	Log HR	HR Estimate	95% CI of HR	P Value
Cox model	N/A	N/A	N/A	N/A
Time-dependent Cox model	3.44	2767	(-6373, 10367)	0.9883
RPSFT model	0.403	1.496	(-139.6, 145.8)	0.9211
Marginal Structural Model	-1.848	0.1574	(0.149, 0.165)	<.0001
Marginal Structural Model (Stabilized Weight)	-1.885	0.1518	(0.143, 0.159)	<.0001
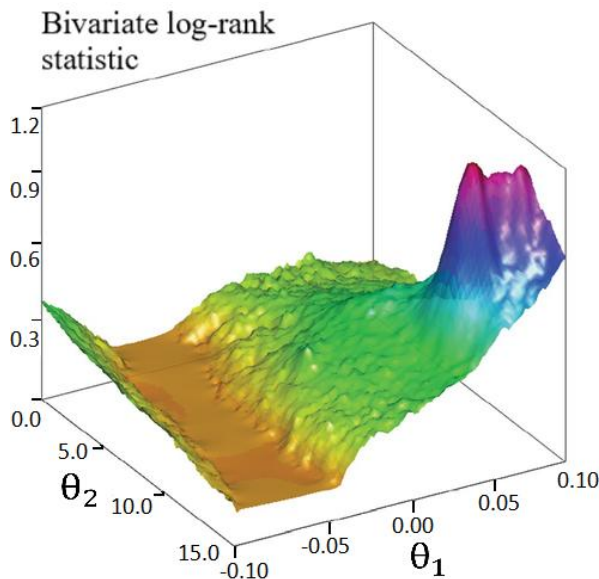
Figure 1.   Surface plot of bivariate log-rank statistic on a grid of (θ_1, θ_2)

5. Conclusions

In this paper, we considered two types of structural models, structural nested failure time (RPSFT) models and marginal structural models, to draw valid causal inference with the presence of secondary salvage treatment. Comparing the results with time-dependent Cox model and marginal structural model, we notice that they both yielded accurate primary treatment effect estimates. However, the implementation of marginal structural model is much more sophisticated than that of time-dependent Cox model. The choice between time-dependent Cox model and marginal structural model depends upon whether the secondary treatment effect is of interest. On the other hand, we should chose marginal structural model when we are interested in the secondary treatment effect.

Furthermore, when an Accelerated Failure Time (AFT) model is preferred, we should chose RPSFT model to obtain valid estimates of primary treatment effect. However, since RPSFT models ignore the confounders that relate to both initiation of secondary treatment and survival time, it usually fails to give accurate secondary treatment effect estimate. An extension of this model [9] was proposed to take care of this issue. Unfortunately, it did not seem to work with another study [10] since the confidence interval for θ_2 was still too wide to be reliable.

In summary, time-dependent Cox model, marginal structural or RPSFT model yield correct primary treatment effect estimate. However, only marginal structural model is able to estimate secondary treatment effect accurately. More work is needed to develop an extension of RPSFT model to provide more reliable estimate of secondary treatment effect.

REFERENCES

[1]   M. Rosenblum, N. P. Jewell, M. van der Laan, S. Shiboski, A. van der Straten, N. Padian. (2009) Analysing direct effects in randomized trials with secondary interventions: an application to human immunodeficiency virus prevention trials. Journal of the Royal Statistical Society. Series A, (Statistics in Society), 172, 443-465.

[2]   M. A. Hernan, B. Brumback and J. M. Robins. (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology, 11(5), 561-570.

[3]   M. A. Hernan, B. Brumback and J. M. Robins. (2001) Marginal structural models to estimate the joint causal effect of nonrandomized treatments. Journal of the American Statistical Association, 96 (454), 440-448.

[4]   M. A. Hernan, S. Cole, J. Margolick, M. Cohen and J. M. Robins. (2005) Structural accelerated failure time models for survival analysis in studies with time-varying treatments. Pharmacoepidemiology and Drug Safety, 14, 477-491.

[5]   M. M. Joffe, W. P. Yang and H. Feldman. (2012) G-estimation and artificial censoring: Problems, challenges, and applications. Biometrics 68 (1): 275-286.

[6]   J. M. Robins, A. A. Tsiatis. (1991) Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics - Theory and Methods 20(8): 2609–2631.

[7]   J. M. Robins and S. Greenland. (1994) Adjusting for differential rates of prophylaxis therapy for PCP in high-versus low-dose AZT treatment arms in an AIDS randomized trial. Journal of the American Statistical Association, 89 (427), 737-749.

[8]   L. J. WEI, Z. YING and D. Y. LIN. (1990) Linear regression analysis of censored survival data based on rank tests. Biometrika, 77(4), 845-851.

[9]   T. Yamaguchi and Y. Ohashi. (2004) Adjusting for differential proportions of second-line treatment in cancer clinical trials. Statistics in Medicine, 23(13), 1991-2003.

[10]   T. Yamaguchi and Y. Ohashi. (2004) Adjusting for differential proportions of second-line treatment in cancer clinical trials. part II: An application in a clinical trial of unresectable non-small-cell lung cancer. Statistics in Medicine, 23 (13): 2005-2022.