

Article Subgroup Identification in Survival Outcome Data Based on Concordance Probability Measurement

Shengli An¹, Peter Zhang² and Hong-Bin Fang^{3,*}

- ¹ Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou 510515, China; hsasl@smu.edu.cn
- ² Otsuka Pharmaceutical Development & Commercialization Inc., Rockville, MD 20878, USA; peter.zhang@otsuka-us.com
- ³ Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington, DC 20057, USA
- * Correspondence: hf183@georgetown.edu

Abstract: Identifying a subgroup of patients who may have an enhanced treatment effect in a randomized clinical trial has received increasing attention recently. For time-to-event outcomes, it is a challenge to define the effectiveness of a treatment and to choose a cutoff time point for identifying subgroup membership, especially in trials in which the two treatment arms do not differ in overall survival. In this paper, we propose a mixture cure model to identify a subgroup for a new treatment that was compared to a classical treatment (or placebo) in a randomized clinical trial with respect to survival time. Using the concordance probability measurement (*K*-index), we propose a statistic to test the existence of subgroups with effective treatments in the treatment arm. Subsequently, the subgroup is defined by a limited number of covariates based on the estimated area under the curve (AUC). The performance of this method in different scenarios is assessed through simulation studies. A real data example is also provided for illustration.

Keywords: concordance probability; *K*-index; mixture models; randomized clinical trials; subgroups; survival analysis

MSC: 62N03; 62G10

1. Introduction

Randomized clinical trials provide conclusive information on treatments by comparing the existing standard treatment with a new treatment, with the aim of increasing the time to failure of the treated patient. Although phase II trials provide sufficient information, it often fails to find an efficient treatment that benefits all patients. However, a new treatment is sometimes shown to be slightly better than the standard therapy, but not sufficiently better for the entire population. In such cases, there may be a subgroup of patients for whom the new treatment provides a substantial benefit. This means that the survival time of a certain subgroup of patients shows a significant improvement under the new treatment, while others may not experience any improvement or may even experience decreased survival in the treatment arm. In such situations, it is challenging to identify such a subgroup and demonstrate that the benefit is likely to be real for future patients [1].

The motivation for this analysis is based on a use case in an AIDS clinical trial, where HIV patients were randomized into the control group receiving monotherapy with didanosine and the treatment group receiving combination therapy with didanosine and zidovudine [2]. In this clinical trial, the effectiveness of the new combination therapy in patients with CD4+ T cells between 200 and 500 mm³ was evaluated by comparing it with the monotherapy treatment for HIV infection. The survival time was defined as the elapsed time from the initial treatment to an event indicating progression to AIDS. A



Citation: An, S.; Zhang, P.; Fang, H.-B. Subgroup Identification in Survival Outcome Data Based on Concordance Probability Measurement. *Mathematics* **2023**, *11*, 2855. https://doi.org/10.3390/ math11132855

Academic Editors: Miguel Ángel Montero-Alonso and Juan De Dios Luna del Castillo

Received: 31 March 2023 Revised: 11 June 2023 Accepted: 24 June 2023 Published: 26 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).



total of 18 demographic and clinical characteristics of patients were considered, and there were no significant differences between the control and treatment groups for any of them (*p*-value > 0.2), indicating that the two arms were balanced. The log-rank test showed that the combination therapy of didanosine and zidovudine did not have a significant difference compared to the monotherapy of didanosine (*p*-value = 0.181, see Figure 1 below), although the combination therapy was slightly better than the monotherapy for the whole population. Despite the balanced covariates between the monotherapy and combination therapy groups, it is interesting to explore if there is a subgroup for whom the combination therapy of didanosine and zidovudine is more effective.



Figure 1. The comparison of combination therapy (treatment) with monotherapy (control).

The search for subgroups with differential treatment effects is known as subgroup identification (see [3–6]). When outcomes from competing treatments are observed for all subjects in both arms, subgroup identification can be achieved using classification or supervised learning algorithms (see, for example, [7–9]). However, in common clinical trials, each patient receives only one treatment, resulting in an imbalance between the new treatment and placebo or an existing treatment for certain subgroups. Consequently, various statistical approaches to subgroup identification have been developed. Lipkovich and Dmitrienko [10] proposed a two-stage procedure in which a small number of biomarkers with the highest predictive ability, based on an appropriate variable importance score, were selected. Subsequently, subgroups with enhanced treatment effects were identified based on the selected biomarkers. Shen and He [11] introduced a structured logistic-normal mixture model for subgroup analysis. Ballarini et al. [12] provided a comparative analysis of different modeling strategies to estimate the predicted individual treatment effect. Additionally, Lipkovich et al. [13] presented a comprehensive review of a broad class of statistical methods used in subgroup identification.

The methods mentioned above for treatment noncompliance in observational data have primarily focused on continuous or binary outcomes and have received relatively less scrutiny with respect to survival outcomes. There are relatively few methods available specifically for identifying subgroups with right-censored survival endpoints (see [14–17]). Loeys and Goetghebeur [18] proposed the structural accelerated failure time (SAFT) models for estimating the causal effects of treatment. Zhang et al. [19] presented a nonparametric method for value function guided subgroup identification. Hu et al. [20] introduced nonparametric Bayesian additive regression trees within the framework of accelerated failure time models. Altstein and Li [21] proposed a semiparametric accelerated failure time time treatment effects in a subgroup of interest with a time-to-event outcome in randomized clinical trials. However, to the best of our knowledge, there are currently no methods available for testing the existence of a subgroup with a significant

treatment effect when the survival time for a new treatment is not sufficiently better than the control group.

In this paper, our objective is to develop a method for identifying a subgroup of patients who respond effectively to a new treatment that may not benefit all patients in a clinical trial. Specifically, we consider a Cox proportional hazards cure model ([22,23]), where patients with an effective treatment response are defined through a latent logistic model. While Wu et al. [24] proposed a likelihood ratio test procedure to assess the existence of subgroups based on the Cox proportional hazards cure model, their testing procedure only determines the presence of a cured subgroup. Since the concordance probability, using the *K*-index has been utilized to measure prognostic accuracy in survival settings ([25,26]), we propose a statistic to test the existence of subgroups with differential treatment effects. Additionally, we aim to identify the specific subgroup based on its association with subject-specific variables.

The rest of the paper is organized as follows. In Section 2, we introduce the Cox proportional hazards cure model for identifying subgroups with effective treatment in a randomized clinical trial, where differential treatment effects exist. We also propose a statistical test procedure using the *K*-index to determine the presence of such subgroups. Section 3 examines the performance of our proposed test through numerical simulations. Furthermore, in Section 4, we apply the proposed methodology to the aforementioned clinical trial, highlighting its potential usefulness. We conclude the paper in Section 5 and provide a brief discussion in Section 6.

2. Methods

2.1. Structured Cure Models

Consider a two-arm clinical trial a random sample of *n* subjects receiving one of the two pre-specified treatments. For patient *i*, we observe $(X_i, \delta_i, \mathbf{Z}_i, TR_i)$ for i = 1, 2, ..., n, where $X_i = \min\{T_i, C_i\}$ is the observed time, and T_i and C_i are the survival and censoring times, respectively. $\delta_i = I(T_i < C_i)$ is the censoring indicator, $\mathbf{Z}_i \in \mathcal{R}^q$ is the observed *q*-dimensional covariates of patient *i*, and TR_i is the treatment indicator with $TR_i = 0$ if patient *i* in the control group and $TR_i = 1$ if patient *i* in the treatment group. Let T_0 be the study end time point, i.e., $X_i \in (0, T_0)$. Usually, the Cox proportional hazards model is employed to patient survival time T_i .

$$\lambda(t|\mathbf{Z}_i) = \lambda_0(t) exp(\mathbf{Z}_i'\beta + TR_i\beta_{TR})$$
(1)

 $t \in \mathcal{R}^+ = (0, \infty)$, for i = 1, 2, ..., n, where β and β_{TR} are unknown coefficients to be estimated. When the estimated β_{TR} is significant negative, the patients in the treatment group have significant treatment effects. In this paper, we want to know if there is a subgroup in the treatment group such that the patients have a significantly longer survival time compared with that in the control group.

For patient *i* in the treatment group $(TR_i = 1)$ with the given covariates Z_i , it is assumed that the longer survival time T_i or X_i indicates patient *i* having the better treatment effect. To investigate the treatment effect, let Y_i be a latent subgroup indicator that is dependent on the covariates Z_i such that $Y_i = 0$ if patient *i* has the treatment effect and otherwise $Y_i = 1$. For patient *i* in the treatment group, the following logistic model is employed for identifying the subgroup with treatment effects,

$$P(Y_i = 1 | \mathbf{Z}_i) = \pi_{\gamma}(\mathbf{Z}_i) = \frac{exp(W'_i\gamma)}{1 + exp(W'_i\gamma)},$$

$$P(Y_i = 0 | \mathbf{Z}_i) = 1 - \pi_{\gamma}(\mathbf{Z}_i) = 1 - P(Y_i = 1 | \mathbf{Z}_i)$$
(2)

where $W_i = (1, Z'_i)'$. Conditional on $Y_i = 1$, the survival time T_i follows the Cox proportional hazards model

$$\lambda_{u}(t|\mathbf{Z}_{i}) = \lambda_{0}(t)exp(\mathbf{Z}_{i}^{\prime}\beta), t \in \mathcal{R}^{+} = (0, \infty),$$
(3)

where γ and β are unknown coefficients to be estimated. Z_i in (2) and (3) may share some, none or all covariates. Then, we have a semiparametric logistic/proportional-hazards mixture model

$$S(t|\mathbf{Z}_i) = \pi_{\gamma}(\mathbf{Z}_i)S_u(t|\mathbf{Z}_i) + 1 - \pi_{\gamma}(\mathbf{Z}_i),$$
(4)

for i = 1, 2, ..., n, where $S_u(t|Z_i)$ is the survival function with its hazard function $\lambda_u(t|Z_i)$ in (3). Unlike the cured models defined in [22,23], the subgroup indicator Y_i is unobserved and dependent on the covariates Z_i which results in more complicates for estimating the parameters in model (4). Since the longer survival time is assumed to be the better treatment effect, we define a new indicator $Y_i(s)$ depending on time *s* to search some subgroups. For a given time $s \in (0, T_0)$, we define

$$Y_i(s) = \begin{cases} 0 & \text{if } T_i > s \quad \text{patient } i \text{ has the treatment effect,} \\ 1 & \text{if } T_i \le s \quad \text{patient } i \text{ has no treatment effect,} \end{cases}$$
(5)

instead of Y_i in model (2). The estimates of the corresponding parameters γ and β which are same as in (2) and (3) can be obtained via the EM algorithm such as in the cured models ([22,23]). Of course, the estimators $\hat{\gamma}$ and $\hat{\beta}$ are dependent on the given time point *s*.

2.2. Hypothesis on Existence of Subgroups

To assess the strength of a risk classification system in survival analysis, the concordance probability is employed as a metric for the global assessment of discrimination. Based on the Cox proportional hazards model, the larger risk score $R_i = \mathbf{Z}'_i \beta$ is, the shorter survival time is. We consider the concordance probability or *K*-index [25]

$$K = P\{T_i > T_j | R_j > R_i, Y_i = Y_j = 1\}.$$
(6)

Obviously, $1/2 \le K \le 1$. K = 1 implies that $R_j > R_i$ can perfectly predict $T_i > T_j$. K = 1/2 is similar to coin tossing, with poor predictive value. The *K*-index is a predictive probability for patients without treatment effects. Based on the definition (6), the *K*-index is dependent on time *t*, denoted by K(t). Let $\theta = (\beta', \gamma')'$. If there is no subgroup for a given time *t*, the parameters γ in model (2) should be zero, and the corresponding index $K(t; \theta)$ is equal to the $K_0(t; \beta_0)$ based on a Cox proportional hazard model

$$\lambda(t|\mathbf{Z}_i) = \lambda_0(t) exp(\mathbf{Z}'_i \beta_0), \quad t \in \mathcal{R}^+ = (0, \infty), \tag{7}$$

with the data $\left\{ \left(\widetilde{X}_i, \ \widetilde{\delta}_i, \mathbf{Z}_i \right); i = 1, \ldots, n \right\},$

$$\left(\widetilde{X}_{i}, \ \widetilde{\delta}_{i}\right) = \begin{cases} (X_{i}, \ \delta_{i}) & \text{if } X_{i} \leq t\\ (X_{i}, \ 0) & \text{if } X_{i} > t. \end{cases}$$

$$(8)$$

Therefore, we propose the following hypothesis for testing the existence of subgroups,

$$H_0: K(t; \boldsymbol{\theta}) \equiv K_0(t; \beta_0), \quad \text{for all } t \in (0, T_0]$$
(9)

and the alternative hypothesis

$$H_1: K(t_0; \boldsymbol{\theta}) \neq K_0(t_0; \beta_0), \quad for \text{ some } t_0 \in (0, T_0].$$

For a given time point *t*, Gonen and Heller [25] proposed to estimate the index $K_0(t; \beta_0)$ as

$$\widehat{K}_{0,n}(t;\hat{\beta}_{0}) = \frac{2}{n(n-1)} \sum_{i \neq j} \frac{I(\mathbf{Z}'_{i}\hat{\beta}_{0} < \mathbf{Z}'_{j}\hat{\beta}_{0})}{1 + \exp(\mathbf{Z}'_{j}\hat{\beta}_{0} - \mathbf{Z}'_{i}\hat{\beta}_{0})},$$
(10)

where $\hat{\beta}_0$ is the estimates of β_0 . $\hat{K}_{0,n}(t; \hat{\beta}_0)$ is a consistent estimate of $K_0(t; \beta_0)$ and has an asymptotically normal distribution which is independent on the unknown censoring

distribution. For the treatment effects defined in (5), let $\hat{\theta} = (\beta', \gamma')'$ be the parametric estimates for the mixture model (4). The K-index based on logistic/proportional hazards mixture model is estimated with marginal probability as

$$\widehat{K}_{n}\left(t;\widehat{\boldsymbol{\theta}}\right) = \frac{\sum_{i\neq j} G_{ij}(\widehat{\boldsymbol{\beta}}) I\left(\mathbf{Z}_{i}'\widehat{\boldsymbol{\beta}} < \mathbf{Z}_{j}'\widehat{\boldsymbol{\beta}}\right) \pi_{\widehat{\gamma}}(\mathbf{Z}_{i}) \pi_{\widehat{\gamma}}(\mathbf{Z}_{j})}{\sum_{i\neq j} I\left(\mathbf{Z}_{i}'\widehat{\boldsymbol{\beta}} < \mathbf{Z}_{j}'\widehat{\boldsymbol{\beta}}\right) \pi_{\widehat{\gamma}}(\mathbf{Z}_{i}) \pi_{\widehat{\gamma}}(\mathbf{Z}_{j})},$$
(11)

where

$$G_{ij}(\hat{eta}) = rac{1}{1 + \exp\left(oldsymbol{Z}_i'\hat{eta} - oldsymbol{Z}_j'\hat{eta}
ight)}$$

The estimator $\hat{K}_n(t; \hat{\theta})$ does not depend on the baseline survival hazards function $\lambda_0(t)$ and is consistent and asymptotically normal under some regularity conditions (Theorem 2 in [26]),

$$\frac{\sqrt{n}\left[\widehat{K}_{n}\left(t;\widehat{\boldsymbol{\theta}}\right)-K(t)\right]}{\sigma_{K}} \to N(0,\ 1),\tag{12}$$

where $\sigma_K^2/n = var(\widehat{K}_n(t;\widehat{\theta})).$

Let $0 < t_{(1)} < t_{(2)} < \ldots < t_{(m)} \le T_0$ be the ordered time points of $\{X_i, i = 1, 2, \ldots, n\}$. To test the hypothesis H_0 in (9), we propose the statistics, for $j = 1, 2, \ldots, m$,

$$D_{j,n} = \frac{\widehat{K}_{0,n}(t_{(j)}; \hat{\beta}_0) - \widehat{K}_n(t_{(j)}; \widehat{\theta})}{\sqrt{\sigma_{\widehat{K}_{0,n}(t_{(j)}; \hat{\beta}_0)}^2 + \sigma_{\widehat{K}_n(t_{(j)}; \widehat{\theta})}^2}} \to N(0, 1),$$
(13)

as $n \to \infty$ where

$$\frac{1}{n}\sigma_{\widehat{K}_{0,n}(t_{(j)};\widehat{\beta}_0)}^2 = var\left\{\widehat{K}_{0,n}\left(t_{(j)};\widehat{\beta}_0\right)\right\}$$

and

$$\sigma_{\widehat{K}_n(t_{(j)};\widehat{\theta})}^2 = nvar\left\{\widehat{K}_n\left(t_{(j)};\widehat{\theta}\right)\right\}.$$

For a given significant level α , define

$$P_n = \min\{1 - \Phi(|D_{j,n}|); j = 1, 2, \dots, m\}.$$
(14)

where $\Phi(\cdot)$ is the distribution function of N(0, 1). If $P_n \ge \alpha/2$, we cannot reject the null hypothesis H_0 . Otherwise, there is some j_0 such that $1 - \Phi(|D_{j,n}|) < \alpha/2$ and there is a subgroup for treatment effects with given time point t_{j_0} in (5).

When the hypothesis H_0 is rejected, we consider the area under the *ROC* curve (*AUC*) for treatment effects with given time point $t_{(j_{cut})}$, which is the first *t* to reject H_0 in ascending survival time order.

$$AUC = P(W'_i \gamma > W'_j \gamma | Y_i = 1, Y_j = 0).$$
(15)

The consistent estimator of AUC is given by [26]

$$\widehat{AUC}\left(\widehat{\boldsymbol{\theta}}, \, \widehat{\lambda}_{0}(t)\right) = \frac{\sum_{i \neq j} I\left(\boldsymbol{W}_{i}^{\prime} \gamma > \boldsymbol{W}_{j}^{\prime} \gamma\right) \widehat{v}_{i}\left(1 - \widehat{v}_{j}\right)}{\sum_{i \neq j} \widehat{v}_{i}\left(1 - \widehat{v}_{j}\right)},\tag{16}$$

where \hat{v}_i is the estimated conditional probability of no treatment effective patients $v_i = P(Y_i = 1 | X \le t_{(j_{cut})}, \delta_i, \mathbf{Z}_i)$. The estimator of v_i is the byproduct of E-step in the

EM algorithm for the mixture model (4). If $\widehat{AUC}(\hat{\theta}, \hat{\lambda}_0(t))$ is significantly greater than 0.5, a subgroup with effective treatment can be found based on the mixture model (4).

3. Simulation Studies

To examine the finite sample performance of the proposed test in (13), this section presents simulation studies. For uncured subjects (without treatment effect), we generated survival time data from the following model

$$\lambda(t|Z_{1i}, Z_{2i}) = \lambda_0(t) exp(\beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_{TR} TR_i),$$
(17)

for $t \in \mathcal{R}^+ = (0, \infty)$, where the baseline survival function is from Weibull distribution with the base hazard function $\lambda_0(t) = 1$. { Z_{1i} , i = 1, ..., n} were generated from $Z_1 \sim N(0, 2^2)$, and { Z_{2i} , i = 1, ..., n} were generated from $Z_2 \sim B(n, p)$ with p = 0.5. We assume that the survival time has no significant difference between the control and treatment groups and set $\beta_{TR} = 0$ in (17). β_1 and β_2 were both set to be log(2). We generated the censoring times from a uniform distribution, U(0,q), where q was chosen to represent the percentage of censored observations in uncured subjects. While we also considered the exponential distribution and other distributions for the censoring times, the results were similar and are not presented in this study.

For the cure status (patients with treatment effects), we chose the following logistic model with covariates W_1 and W_2 which were generated in the same way as Z_1 and Z_2 :

$$\pi_b(Z_i) = \frac{exp(b_0 + b_1 Z_{1i} + b_2 Z_{2i})}{1 + exp(b_0 + b_1 Z_{1i} + b_2 Z_{2i})},$$
(18)

Both b_1 and b_2 were set to be log(2), and b_0 was selected for the percentages of cured observations in each replication. The uncured status *Y* was assumed to follow a Bernoulli distribution with a success probability π_b in (18). Without loss of generality, the simulation studies considered the case where the censored rate exceeds the cured rate. As the cured rates increased, the testing power also increased. Two sample sizes, n = 200 and 400, were used for the simulations. For each simulated dataset, the D_n statistic was calculated using Formula (13) and H_0 was tested using Formula (14). The procedure was repeated 1000 times to assess the type I errors and powers of our proposed method.

When there are no cured individuals (i.e., no subgroups with treatment effects), Table 1 demonstrates that the type I error increases with ascending censored proportions. However, all of the type I errors are still able to maintain the nominal significant level of 5%.

Table 1. Type I error under different scenarios.

Censored Proportion				
n	0.10	0.30	0.50	
200	0.001	0.002	0.012	
400	0.004	0.008	0.049	

n: sample size; simulation times = 1000.

Conditional on a fixed cured proportion (e.g., 30%), the power will decrease as the censored rates increase. However, even with increasing censored rates, the power remains sufficiently high (e.g., 85%) as long as the differences between the cured and censored proportions are <15% with a sample size of n = 200. With a larger sample size of n = 400, the power can still be maintained at >80% even with higher censored rates (see Table 2). When the cured proportion is 50%, it can be observed that the power remains high only when the sample size is large (e.g., n = 400) and the differences between the cured and censored proportions are <15%.

	Cured Proportion (Censored Proportion)				
п	0.1 (0.2)	0.2 (0.3)	0.3 (0.4)	0.4 (0.5)	0.5 (0.6)
200	0.909	0.952	0.918	0.864	0.668
400	0.995	0.999	0.999	0.984	0.934

Table 2. Power analysis with different scenarios.

n: sample size; simulation times = 1000.

With a fixed difference between cured and censored proportions (e.g., 10%), the power decreases as the cured proportions increase. However, if the sample size is sufficiently large (e.g., n = 400), the power can remain high (>90%). The simulation results are summarized in Table 3.

Table 3. Power with fixed differences between cured and censored proportions.

		Censored Proportion			
п	cured rate	0.35	0.40	0.45	0.50
200	0.30	0.968	0.929	0.850	0.696
400	0.30	1.000	0.999	0.989	0.955
			Censored	Proportion	
п	cured rate	0.55	0.60	0.65	0.70
200	0.50	0.812	0.661	0.475	0.293
400	0.50	0.981	0.923	0.809	0.571

Censored proportion—cured proportion = 0.10.

With our proposed method, we first determine the time to cure $t_{(j_{cut})}$, and then obtain predicted subgroup memberships and their accuracy (the proportion of correctly classified). A logistic model can be constructed using the predicted subgroups and associated covariates for further subgroup identification. Table 4 presents the accuracy and area under the curve (AUC) obtained from logistic models, which remain adequate across most scenarios. A higher censored proportion results in lower accuracy and given a fixed difference between the cured and censored proportions, the accuracy decreases as the cured proportion increases.

Table 4. Accuracy of the proposed method.

Cured Proportion	Censored Proportion	РСС	AUC
0.3	0.4	0.813	0.779
0.3	0.5	0.727	0.737
0.5	0.6	0.788	0.774
0.5	0.7	0.681	0.703

PCC: proportion of classified correctly.

4. Application to Clinical Trial Data

We analyzed the real data ACTG175 [2] with our proposed method. In this study, the sample sizes of "combination therapy of zidovudine and didanosine" and "monotherapy didanosine" are $n_{tr} = 522$ and $n_{con} = 559$, respectively. The test showed subgroups may exist ($D_{j_0, n} = 2.364$, *p*-value = 0.009) in the treatment group with the time $t_{j0} = day 486$. The model to identify subgroups is $logist(P(Y = 1)) = -3.5506 + 0.5783 \times symptom + 2.4904 \times offtrt, which shows that patients' symptom (symptomatic indicator, <math>0 = asymptomatic$, 1 = symptomatic) and offtrt (indicator of off-treatment before 96 ± 5 weeks, 0 = no, 1 = yes). The *AUC* based on the logistic model is 0.7903 (95% CI: (0.7342, 0.8465)) and the best.

The probability threshold was 0.1530 with the corresponding sensitivity and specificity at 81.7% and 72.9%. The log-rank test shows that the treatment effect of the beneficial

subgroup Y = 0 is better than the control group (*p*-value = 0.000253), which is different from the result of the total treatment group vs. the control group (*p*-value = 0.181). Within the treatment group (combination therapy of zidovudine and didanosine), the beneficial subgroup Y = 0 also showed a better effect than the non-beneficial subgroup Y = 1 (*p*-value < 0.0001, see Figure 2).



Figure 2. The comparison of two subgroups (combination therapy) with monotherapy (control), Y = 0 means the beneficial subgroup, and Y = 1 means otherwise.

5. Conclusions

In clinical trials, it is recognized that certain patients may respond differently to treatment based on their individual characteristics. Existing methods for identifying effective subgroups with time-to-event outcomes often rely on estimating the survival function for the overall population using a predefined cutoff time point as the maximum observed follow-up time (e.g., [12,19]). However, since patients with different biomedical characteristics may exhibit varying treatment efficacies in the early and later time periods, this paper proposes a testing procedure for identifying potential subgroups using survival function estimates at different cutoff time points with right-censored time-to-event data.

The proposed test utilizes two types of *K*-indexes based on the semiparametric Logistic– Cox mixture model. Through simulation studies and real data analysis, it is demonstrated that the proposed approach effectively detects subgroups and identifies the time point at which survival statuses differ. These methods can also be adapted to other fields such as cancer prevention, public management, and marketing, where a particular public policy or marketing strategy may have varying effects on different groups of individuals. The results of this study contribute to the understanding of subgroup-specific treatment effects and provide a valuable tool for tailoring interventions to specific patient groups.

6. Discussions

There has been a recent surge in the development of methods for subgroup identification, recognizing the importance of associated covariates in this process. It is valuable to explore the role of covariates in identifying subgroups and to construct models for predicting subgroup identification in practical applications. Our research focuses on investigating effective semiparametric models in causal inference with time-dependent outcomes, such as commonly encountered survival data and dynamic treatment regimes.

However, it is essential to note that our proposed method relies on the correct specification of the assumed statistical models. Specifically, the modeling assumptions of the Cox proportional hazards framework must be met for the survival component of the mixture model. It is crucial to identify and employ model diagnostic procedures to ensure the validity of the proposed method. Furthermore, our method requires a relatively large sample size and relies on the differences between the cured and censored proportions.

In observational epidemiological studies or clinical trials without completely random treatment assignments, naive estimators of treatment effects based on treatment group data can be biased due to confounding. Consequently, causal inference methods are necessary to obtain unbiased estimators and establish causal relationships ([27]). These challenges and limitations present opportunities for further study and improvement in the future. Overall, addressing these issues and advancing the field of causal inference in the context of time-dependent outcomes will contribute to the development of more robust and reliable methods for subgroup identification and causal inference in clinical and epidemiological research.

Author Contributions: Conceptualization, P.Z. and H.-B.F.; methodology, S.A. and H.-B.F.; validation, P.Z.; formal analysis, S.A. and P.Z.; writing—original draft preparation, S.A.; and writing—review and editing, H.-B.F. and P.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The research of An and Fang is partially supported by the National Cancer Institute (NCI) grant R01CA164717 and P30CA051008. An has conducted his work at Georgetown University when he was as a visiting scholar partially supported by China State Scholarship Foundation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

- Ruberg, S.; Chen, L.; Wang, Y. The Mean Does Not Mean as Much Anymore: Finding Sub-Groups for Tailored Therapeutics. *Clin. Trials* 2010, 7, 574–583. [CrossRef]
- Hammer, S.; Katzenstein, D.; Hughes, M.; Gundacker, H.; Schooley, R.; Haubrich, R.; Henry, W.; Lederman, M.; Phair, J.; Niu, M.; et al. A Trial Comparing Nucleoside Monotherapy with Combination Therapy in HIV-Infected Adults with CD4 Cell Counts from 200 to 500 per Cubic Millimeter. N. Engl. J. Med. 1996, 335, 1081–1090. [CrossRef] [PubMed]
- Follmann, D. On the Effect of Treatment among Would-Be Treatment Compliers: An Analysis of the Multiple Risk Factor Intervention Trial. J. Am. Stat. Assoc. 2000, 95, 1101–1109. [CrossRef]
- Foster, J.; Taylor, J.; Ruberg, S. Subgroup Identification from Randomized Clinical Trial Data. *Stat. Med.* 2011, 30, 2867–2880. [CrossRef]
- Lipkovich, I.; Dmitrienko, A.; Denne, J.; Enas, G. Subgroup Identification Based on Differential Effect Search-A Recursive Partitioning Method for Establishing Response to Treatment in Patient Subpopulations. *Stat. Med.* 2011, 30, 2601–2621. [CrossRef]
- 6. Zhao, L.; Tian, L.; Cai, T.; Claggett, B.; Wei, L. Effectively Selecting a Target Population for a Future Comparative Study. J. Am. Stat. Assoc. 2013, 108, 527–539. [CrossRef]
- 7. Bonetti, M.; Gelber, R. Patterns of Treatment Effects in Subsets of Patients in Clinical Trials. Biostatistics 2004, 5, 465–481. [CrossRef]
- 8. Xiao, S.; Pepe, M. Evaluating Markers for Selecting a Patient's Treatment. Biometrics 2004, 60, 874–883. [CrossRef]
- Loh, W.; He, X.; Man, M. A Regression Tree Approach to Identifying Subgroups with Differential Treatment Effects. *Stat. Med.* 2015, 34, 1818–1833. [CrossRef] [PubMed]
- 10. Lipkovich, I.; Dmitrienko, A. Strategies for identifying predictive biomarkers and subgroups with enhanced treatment effect in clinical trials using SIDES. *J. Biopharm. Stat.* **2014**, *24*, 130–153. [CrossRef] [PubMed]
- 11. Shen, J.; He, X. Inference for Subgroup Analysis with a Structured Logistic-Normal Mixture Model. J. Am. Stat. Assoc. 2015, 110, 303–312. [CrossRef]
- 12. Ballarini, N.M.; Rosenkranz, G.K.; Jaki, T.; König, F.; Posch, M. Subgroup identification in clinical trials via the predicted individual treatment effect. *PLoS ONE* **2018**, *13*, e0205971. [CrossRef]

- Lipkovich, I.; Dmitrienko, A.; D'Agostino Sr, R.B. Tutorial in biostatistics: Data-driven subgroup identification and analysis in clinical trials. *Stat. Med.* 2017, 36, 136–196. [CrossRef]
- Ciampi, A.; Negassa, A.; Lou, Z. Tree-structured Prediction for Censored Survival-data and the Cox model. J. Clin. Epidemiol. 1995, 48, 675–689. [CrossRef] [PubMed]
- 15. Goetghebeur, E.; Molenberghs, G. Causal Inference in a Placebo-Controlled Clinical Trial with Binary Outcome and Ordered Compliance. J. Am. Stat. Assoc. 1996, 91, 928–934. [CrossRef]
- 16. Kehl, V.; Ulm, K. Responder Identification in Clinical Trials with Censored Data. *Comput. Stat. Data Anal.* **2006**, *50*, 1338–1355. [CrossRef]
- Cuzick, J.; Sasieni, P.; Myles, J.; Tyrer, J. Estimating the Effect of Treatment in a Proportional Hazards Model in the Presence of Non-Compliance and Contamination. J. R. Stat. Soc. Ser. B 2007, 69, 565–588. [CrossRef]
- 18. Loeys, T.; Goetghebeur, E. A Causal Proportional Hazards Estimator for the Effect of Treatment Actually Received in a Randomized Trial with All-or-Nothing Compliance. *Biometrics* 2003, *59*, 100–105. [CrossRef] [PubMed]
- Zhang, P.; Ma, J.; Chen, X.; Shentu, Y. A nonparametric method for value function guided subgroup identification via gradient tree boosting for censored survival data. *Stat. Med.* 2020, *39*, 4133–4146. [CrossRef] [PubMed]
- Hu, L.; Ji, J.; Li, F. Estimating heterogeneous survival treatment effect in observational data using machine learning. *Stat. Med.* 2021, 40, 4691–4713. [CrossRef]
- Altstein, L.; Li, G. Latent Subgroup Analysis of a Randomized Clinical Trial through a Semiparametric Accelerated Failure Time Mixture Model. *Biometrics* 2013, 69, 52–61. [CrossRef] [PubMed]
- 22. Sy, J.; Taylor, J. Estimation in a Cox Proportional Hazards Cure Model. Biometrics 2000, 56, 227–236. [CrossRef] [PubMed]
- 23. Fang, H.B.; Li, G.; Sun, J. Maximum Likelihood Estimation in a Semiparametric Logistic/Proportional-Hazards Mixture Model. *Scand. J. Stat.* 2005, 32, 59–75. [CrossRef]
- 24. Wu, R.; Zheng, M.; Yu, W. Subgroup Analysis with Time-to-Event Data under a Logistic-Cox Mixture Model. *Scand. J. Stat.* 2016, 43, 863–878. [CrossRef]
- 25. Gonen, M.; Heller, G. Concordance Probability and Discriminatory Power in Proportional Hazards Regression. *Biometrika* 2005, 92, 965–970. [CrossRef]
- Zhang, Y.; Shao, Y. Concordance Measure and Discriminatory Accuracy in Transformation Cure Models. *Biostatistics* 2018, 19, 14–26. [CrossRef]
- 27. Bai, X.; Tsiatis, A.A.; O'Brien, S.M. Doubly robust estimators of treatment-specific survival distributions in observation studies with stratified sampling. *Biometrics* **2013**, *69*, 830–839. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.