JOURNAL OF BIOPHARMACEUTICAL STATISTICS
2021, VOL. 31, NO. 2, 141–155
https://doi.org/10.1080/10543406.2020.1814799

Estimation on conditional restricted mean survival time with counting process

Junshan Qiu[a], Dali Zhou[b], H.M. Jim Hung[c], John Lawrence[b], and Steven Bai[b]

[a]Division of Pharmacometrics, OCP/OTS/CDER, US FDA; [b]Division of Biometrics II, OB/OTS/CDER, US FDA; [c]Division of Biometrics I, OB/OTS/CDER, US FDA

ABSTRACT

In a comparative longitudinal clinical study, multiple clinical events of interest are typically collected in timing and occurrence during the follow-up period. These clinical events are often indicative of disease burden over the study period and provide overall evidence of benefit/risk of one treatment relative to another. While these clinical events are usually used to form a composite endpoint, only the first occurrence of the composite endpoint event is considered in primary efficacy analysis. This type of analysis is commonly performed but it may not be ideal. Most of the existing methods for analyzing multiple event-time data were developed, relying on certain model assumptions. However, the assumptions may greatly affect the inferences for treatment effect. In this paper, we propose a simple, nonparametric estimator of conditional mean survival time for multiple events to quantify treatment effect which has clinically meaningful interpretation. We use simulation studies to evaluate the performance of the new method. Further, we apply this method to analyze the data from a cardiovascular clinical trial as an illustration.

ARTICLE HISTORY
Received 12 February 2020
Accepted 24 July 2020

KEYWORDS
Clinical trials; composite endpoint; multiple events; restricted mean survival time

1. Introduction

It is commonly seen in a longitudinal clinical study that multiple clinical events can occur to each patient at various time points during the follow-up period. A time profile for each patient can be pictured from these clinical events. How to utilize the data is mainly driven by the study objectives, for instance, whether a new therapy is superior to a standard care with respect to benefit and risk. Time-to-first event analysis is commonly considered. However, it may be an efficient way to utilize all relevant data as illustrated in the following example. In history, models and methods based on counts were developed for the scenario that individuals frequently experience the event of interest. The events could be incidental such as mild epileptic seizures or asthmatic attacks in humans or non-incidental including Myocardial Infarction (MI) and stroke in cardiovascular studies in the sense that the occurred events may substantially alter patients' disease status and affect the event process in the future. Event counts are analyzed under the framework of Poisson process initially. To accommodate between-subject variability in event rates via fixed or time-varying covariates or random effects, dramatic extension work to the Poisson process has been done such as intense-based models for multitype recurrent events (Andersen et al. 2012). Negative Binomial (NB) models which can handle random effects and LWYY method (Lin et al. 2000) which can provide robust estimation of marginal features such as rate and mean function. European Medicines Agency (EMA) evaluated several statistical methods including the NB and LWXY

CONTACT Junshan Qiu Junshan.qiu@fda.hhs.gov Division of Pharmacometrics, OCP/OTS/CDER, US FDA
This article reflects the views of the authors and should not be construed to represent FDA's views or policies.[†]Food and Drug Administration, CDER/OTS/OCP/DPM[‡]Food and Drug Administration, CDER/OTS/OB/DBII[§]Food and Drug Administration, CDER/OTS/OB/DBI

in handling recurrent events in the presence of a terminal event with or without informative discontinuations (Hougaard 2019). The performance of these methods is affected by the follow-up time by design and the effect of the terminal event. Further research in this area is needed to address the issues.

PEGASUS-TIMI 54 (Bonaca, 2015)(Bonaca et al. 2015) is a randomized, double-blind, placebo-controlled clinical trial conducted to test the hypothesis that long-term therapy with ticagrelor added to low-dose aspirin reduces the risk of major adverse cardiovascular events among stable patients with a history of MI. A total of 21,162 patients were randomized in a double-blind 1:1:1 fashion to ticagrelor at a dose of 90 or 60 mg twice daily, or placebo. All patients were to receive low-dose aspirin and were followed for a median of 33 months. The primary efficacy endpoint was the composite of cardiovascular death (CVD), MI, or stroke. As shown in Figure 1, the long-term treatment with ticagrelor 90 or 60 mg twice daily in combination with aspirin in patients with history of MI and at high risk of an atherothrombotic event demonstrated a statistically significant benefit in reducing the event rate of the primary composite endpoint [hazard ratio (HR) = 0.85, 95% CI (0.75, 0.96), $p = 0.0080$ for ticagrelor 90 mg; HR = 0.84, 95% CI (0.74, 0.95), $p = 0.0043$ for ticagrelor 60 mg]. Besides the composite endpoint, subsequent analyses are almost surely to look into the distribution of each specific component event. Apparently, CVD events as a component of the composite endpoint are less counted than as one type of multiple events (See upper and lower panels in Table 1), since they can be masked by the earlier events. While MI and stroke events are almost equally counted either as a component of the composite or as one type of multiple events (Table 1). Clearly, the treatment of 90 or 60 mg ticagrelor tends to reduce the occurrence of CVD more if we consider the CVD events alone.

There are several model-based procedures for analyzing such time-to-multiple-event data, see Claggett et al. (2018) (Claggett et al. 2018). To ease the interpretation of the results, Claggett et al. (2018) (Claggett et al. 2018) developed a non-parametric approach using reverse counting process to accommodate a terminal event such as death. Cumulative difference/ratio in total event-free survival time was used as an estimator of treatment effect utilizing both the occurrence and the timing of the events. In this paper, following Qiu et al., 2018 (Qiu et al. 2019), we extend the estimator of conditional restricted mean survival time using the reverse counting process to analyze the aforementioned time-to-multiple-event data with terminal events.

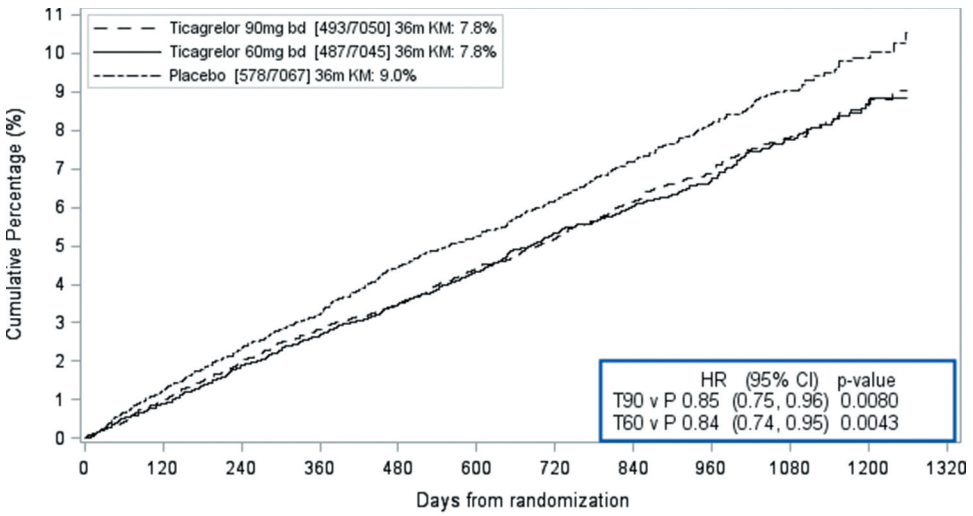

Figure 1. Kaplan–Meier plot of primary efficacy endpoint.

Table 1. Summary of a number of patients experiencing each type of events, and specific type within the composite, by treatment group.

Event type	Placebo	Ticagrelor 90 mg	Ticagrelor 60 mg
Composite and its decomposition			
Composite	578 (8.2%)	493 (7.0%)	487 (6.9%)
CVD	128 (1.8%)	127 (1.8%)	116 (1.6%)
MI	336 (4.8%)	272 (3.9%)	283 (4.0%)
Stroke	114 (1.6%)	94 (1.3%)	88 (1.2%)
Multiple events by type			
CVD	210 (3.0%)	182 (2.6%)	174 (2.5%)
MI	338 (4.8%)	275 (3.9%)	285 (4.0%)
Stroke	122 (1.7%)	100 (1.4%)	91 (1.3%)

1.1. Toy example

We use a toy example to illustrate the utility of our proposed method. First, consider the extreme scenario that there is only one subject per treatment. Suppose that two patients receiving different treatments have the first non-terminal event and the death event occurred at the same time. But one patient who received a treatment has no event between the first event and death, while the other patient on the other group has three events after the first event and before death (Figure 2). Based on either the time-to-first event or time-to-death analysis, these two patients share the same survival time. However, using our proposed approach, the difference in terms of event-free time between the two patients can be evaluated as the area highlighted with stars (Figure 3). The y-axis in Figures 2 or 3 is labelled as 'average event-free probability' which is defined as the average of individual event-probability, which is $1 - (\text{number of events occurred})/(\text{total number of potential events})$ by each time point. For a single patient, it is equivalent to individual event-free probability.

2. Conditional restricted mean survival time for multiple events

Assume a sequence of (K + 1) distinct types of events can be potentially observed for each subject during the follow-up period in a longitudinal clinical trial. Let T_k, $k = 1, \ldots, K+1$, be the minimum of the first occurrence of the kth type of non-terminal event ($k = 1, \ldots, K$) and the terminal event

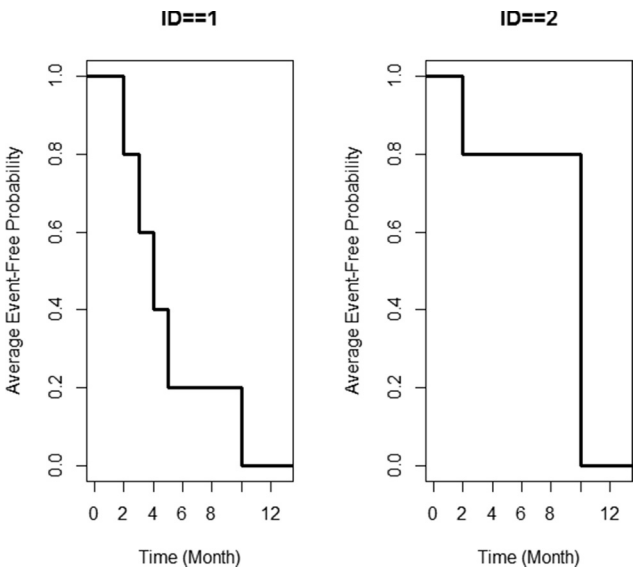

Figure 2. Survival curves for Patient 1 vs. Patient 2.

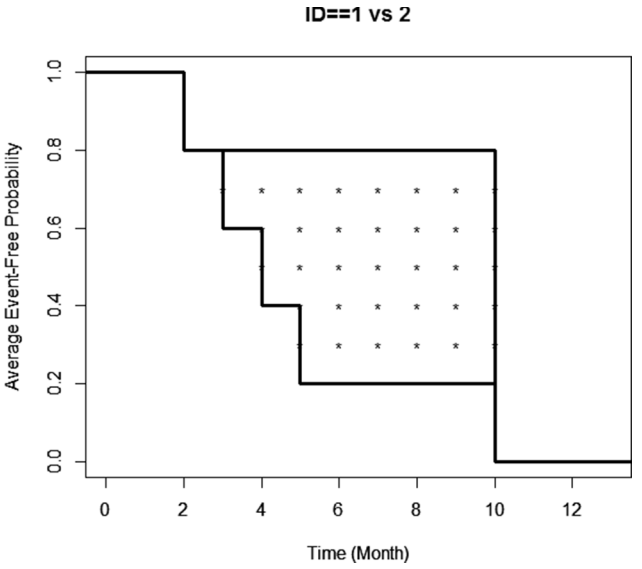

Figure 3. Survival curves for Patient 1 vs. Patient 2.

$(k = K + 1)$. These event times can be censored by a random variable C, which is assumed to be independent of T_k. Let X_k be the minimum of T_k and C. Let Δ_k be the censoring status indicator, that is, $\Delta_k = 1$ if T_k is observed and zero otherwise.

A traditional conditional restricted mean survival time (CRMST) of T with the cut-off time u is defined as the mean of restricted survival time $min(T, u)$ conditional on time interval $[t_1, t_2)$,

$$m(u) = E(min(T, u)|t_1 \leq T < t_2),$$

which can be estimated by the area under the survival curve $S(t)$ over the time interval of interest. Note that, due to the right censoring, u has to be restricted to the interval $[0, \tau]$, where τ satisfies $\prod_{k=1}^{K+1} P(X_k > \tau) = 0$ (Zhao et al., 2016 (Zhao et al. 2016)).

2.1. Estimator of CRMST for multiple events

Let $S_k(\cdot)$ be the survival function for T_k conditional on the interval $[t_1, t_2)$, and $\hat{S}_k(\cdot)$ be the KM estimate of $S_k(\cdot)$. By Qiu et al. (2019), for a time interval $[t_1, t_2)$ of interest on a non-negative domain, the corresponding CRMST for type k event can be estimated by

$$\hat{m}^k_{t_1,t_2}(u) = \frac{1}{n_k} \frac{1}{\hat{S}_k(t_1) - \hat{S}_k(t_2)} \sum_{i=1}^{n_k} \frac{\hat{S}_k(X_{ik-1})\Delta_{ik} I(t_1 \leq X_{ik} < t_2)}{\bar{R}(X_{ik-1})} \int_0^u I(X_{ik} \geq s) ds. \tag{1}$$

For multiple-event case, we combine the CRMST estimators for all $(K + 1)$ types of events and propose the multiple-event CRMST (MCRMST) estimator,

$$\hat{m}_{t_1,t_2}(u) = \sum_{k=1}^{K+1} \hat{m}^k_{t_1,t_2}(u), \tag{2}$$

where $\hat{m}^k_{t_1,t_2}(u)$ is defined in Equation (1). As can be seen from Equation (2), the MCRMST estimator is the summation of CRMST estimators across the $(K + 1)$ types of events. Note that the sample sizes n_k for different k, $k = 1, 2, \ldots, K + 1$ in the above equations are not necessarily the same. Equations (1) and (2) can be further simplified in computation (See Appendix).

2.2. Variance estimation

In this section, we discuss the asymptotic distribution of our estimator, variance estimation using perturbation re-sampling method, hypothesis test, and confidence interval construction.

2.2.1. One-arm case

For the kth type of event, CRMST $m^k_{t_1,t_2}(u)$ and its estimator $\hat{m}^k_{t_1,t_2}(u)$ are the areas under $S_k(\cdot)$ and $\hat{S}_k(\cdot)$ up to u conditional on the time interval $[t_1, t_2)$, respectively, calculated horizontally (Qiu, et al., 2019 (Qiu et al. 2019)). According to Zhao, et al. (2016) (Zhao et al. 2016), the uniform consistency of the KM estimator (Gill, 1983 (Gill 1983)) implies the uniform consistency of RMST estimator. Similarly, for the kth type of event conditional on $[t_1, t_2)$, $\hat{m}^k_{t_1,t_2}(u)$ is uniformly consistent for $m^k_{t_1,t_2}(u)$. Thus, the distribution of $\sqrt{n_k}\{\hat{m}^k_{t_1,t_2}(u) - m^k_{t_1,t_2}(u)\}$ can be approximated by the distribution of a Gaussian process $G(\cdot)$ with mean 0, asymptotically. The limiting distribution can be approximated by perturbation re-sampling method (Claggett et al., 2018 (Claggett et al. 2018); Zhao et al., 2016 (Zhao et al. 2016); Lin et al., 1993 (Lin et al. 1993); Parzen et al., 1997 (Michael Parzen and Wei 1997); Zhao et al., 2012 (Zhao et al. 2012)) when the sample size n_k is big enough. Let n_M be the total sample size of K + 1 types (i.e., $n_M = \sum_{k=1}^{K+1} n_k$). Since our proposed MCRMST estimator $\hat{m}_{t_1,t_2}(u)$ is the summation of $\hat{m}^k_{t_1,t_2}(u)$, therefore the distribution of $\sqrt{n_M}\{\hat{m}_{t_1,t_2}(u) - m_{t_1,t_2}(u)\}$ can be approximated by $G_M(\cdot)$, the summation of (K + 1) Gaussian processes asymptotically, which is also a Gaussian process with mean 0. Denote the standard deviation estimate for the distribution of $G_M(\cdot)$ by $\hat{\sigma}(\cdot)$. Then for each u, the $(1-\alpha)$ confidence interval of $m_{t_1,t_2}(u)$ is

$$(\hat{m}_{t_1,t_2}(u) - z_{(1-\alpha/2)} \cdot n_M^{-1/2}\hat{\sigma}(u), \hat{m}_{t_1,t_2}(u) + z_{(1-\alpha/2)} \cdot n_M^{-1/2}\hat{\sigma}(u)),$$

where z follows standard normal distribution, and the confidence interval can be computationally obtained from perturbation re-sampling. A perturbed version of MCRMST estimator is as follows:

$$\hat{m}^*_{t_1,t_2}(u) = \sum_{k=1}^{K+1} \frac{1}{n_k} \frac{1}{\hat{S}_k(t_1) - \hat{S}_k(t_2)} \sum_{i=1}^{n_k} V_{ik} \frac{\hat{S}_k(X_{ik-1})\Delta_{ik} I(t_1 \le X_{ik} < t_2)}{\bar{R}(X_{ik-1})} \int_0^u I(X_{ik} \ge s) ds, \tag{3}$$

where $\{V_{ik} : i = 1, \ldots, n_k\}$ is the perturbation variable for the kth type of event, generated from standard exponential distribution which is independent of the data. Then, with M (e.g., M=1000) realizations of every perturbation variable $V_{ik}, k = 1, \ldots, K+1$, the distribution of $\sqrt{n_M}\{\hat{m}_{t_1,t_2}(u) - m_{t_1,t_2}(u)\}$ can be approximated by the empirical distribution of $\sqrt{n_M}\{\hat{m}^*_{t_1,t_2}(u) - \hat{m}_{t_1,t_2}(u)\}$.

2.2.2. Two-arm case

Suppose we would like to test whether there is a significant MCRMST difference between two treatment arms: arm 1 and arm 2. Let n_{M1}, n_{M2} be the arm 1 and arm 2 versions of n_M, X_k^{arm1} and X_k^{arm2} be the arm 1 and arm 2 versions of X_k. We consider the hypothesis test,

$$H_0 : D_{t_1,t_2}(u) = 0 vs. H_1 : D_{t_1,t_2}(u) \neq 0,$$

where $D_{t_1,t_2}(u) = m^{arm1}_{t_1,t_2}(u) - m^{arm2}_{t_1,t_2}(u)$ is the MCRMST difference. According to the one-arm case, it follows that $\hat{D}_{t_1,t_2}(u) = \hat{m}^{arm1}_{t_1,t_2}(u) - \hat{m}^{arm2}_{t_1,t_2}(u)$ is a consistent estimator of $D_{t_1,t_2}(u)$ for $u < \tau$, where τ satisfies $\prod_{k=1}^{K+1} P(X_k^{arm1} > \tau) P(X_k^{arm2} > \tau) = 0$. Furthermore, the distribution of $\sqrt{n_{M1} + n_{M2}}(\hat{D}_{t_1,t_2}(u) - D_{t_1,t_2}(u))$ can be approximated by a Gaussian Process. Denote the standard deviation estimate for this Gaussian Process by $\hat{\sigma}^2_D$, and $\hat{\sigma}^2_D = \frac{\hat{\sigma}^2_1}{n_{M1}} + \frac{\hat{\sigma}^2_2}{n_{M2}}$. Then for each u, we construct the following test statistic

$$TS = \frac{\hat{D}_{t_1,t_2}(u)}{\hat{\sigma}_D},$$

of which the variances can be calculated using the similar perturbation re-sampling method used for one-arm case. And based on the Gaussian approximation, TS has a standard normal distribution asymptotically under the null hypothesis.

3. Simulation studies

Simulation studies are conducted to evaluate the statistical performances (biases, coverage probabilities, etc.) of MCRMST at different scenarios: one arm without frailty, one arm with frailty, two arms without frailty, and two arms with frailty. For one-arm cases, we calculate the confidence intervals and compare the coverage probabilities to the nominal level, 95%. For cases with two arms, we perform the hypothesis test proposed in section 2 and present the type-I error and the power under the null and alternative hypotheses, respectively. The nominal significance level is 5%.

The number of subjects n is 400. Assume each patient potentially has up to four types of events including the terminal event, i.e., K + 1 = 4. Event times $T_{i1}, T_{i2}, T_{i3}, T_{i4}$ are generated from Weibull distributions with or without common frailty. The parameter values are specified in the subsections. In order to capture both the random censoring and administrative censoring, the censoring time C is a mixture of uniform distribution and a fixed number (longest follow-up time). Once the terminal event T_{i4} occurs, other types of events occurred after the terminal event is censored. This mimics a quasi-competing risk scenario.

Simulation procedure:

(1) Initialize parameter values t_1, t_2 and u;
(2) To mimic the population, generate event times for each subject from Weibull distribution with a number of subjects $N \geq 1,000,000$. Calculate the 'true' MCRMST by averaging the generated survival times conditional on the time interval (t_1, t_2), empirically;
(3) Repeat the following steps for 1000 times:

- Take a sample from the population with a sample size 400;
- Calculate MCRMST estimator;
- Estimate the variance of MCRMST by performing the perturbation re-sampling method with 4000 realizations;
- Construct the confidence interval and record the coverage indicator (whether it covers the truth or not);
  (1) Average the 1000 repetitions.

3.1. One arm without frailty

Starting from simplest setup, we explore the statistical properties of our estimator in the one-arm case. Simulate event times from Weibull distributions without frailty, with shape parameter 0.8 and scale parameters 1000, 1500, 2000, and 10,000, respectively, for each type of event. All four events are independently generated, there is no correlation between events in this case.

In Table 2, 'nEvn' column provides a number of events observed up to cut-off time 'u'. 'CR' is the average censoring rate of all four events, while 'DECR' is the death event censoring rate. 'Truth' shows empirical true MCRMST of the population. 'RB(%)' is the relative bias of MCRMST estimator with respect to the 'Truth' in percentage. 'SE' shows the perturbation re-sampling standard error of the MCRMST estimator. 'Coverage', coverage probability, is the percentage of 'Truth' covered by the confidence intervals in the 1000 repetitions.

Table 2. MCRMST: one arm without frailty.

t1	u	t2	nEvn	CR	DECR	Truth	Estimator	RB(%)	SE	Coverage
0	600	2930	453.99	0.55	0.81	1936.13	1936.06	−0.00	18.66	0.936
0	800	2930	520.24	0.55	0.81	2452.77	2455.19	0.10	27.52	0.944
0	1000	2930	573.59	0.55	0.81	2925.13	2,930.40	0.18	37.97	0.958
0	1200	2930	613.46	0.55	0.81	3359.85	3367.11	0.22	48.89	0.969
0	1400	2930	643.95	0.55	0.81	3762.19	3771.35	0.24	60.58	0.972
0	1600	2930	667.53	0.55	0.81	4136.13	4146.52	0.25	73.12	0.978
600	800	2930	66.86	0.67	0.81	3158.29	3159.12	0.03	5.37	0.935
600	1000	2930	119.93	0.67	0.81	3843.68	3846.77	0.08	15.64	0.966
600	1200	2930	159.92	0.67	0.81	4467.50	4471.64	0.09	27.95	0.971
600	1400	2930	190.20	0.67	0.81	5039.28	5044.75	0.11	42.10	0.973
600	1600	2930	213.73	0.67	0.81	5566.06	5571.66	0.10	57.78	0.975
600	1800	2930	229.47	0.67	0.81	6053.81	6062.31	0.14	73.32	0.981
1200	1400	2930	30.55	0.76	0.81	5563.22	5565.93	0.05	7.78	0.943
1200	1600	2930	53.97	0.76	0.81	6261.69	6265.14	0.06	20.87	0.953
1200	1800	2930	69.80	0.76	0.81	6904.10	6912.89	0.13	36.65	0.943
1200	2000	2930	84.93	0.76	0.81	7496.28	7512.32	0.21	60.96	0.971
1200	2200	2930	95.26	0.76	0.81	8044.07	8067.43	0.29	85.15	0.975

Notes. nEvn: number of events. CR: censor rate. DECR: death event censor rate. Truth: true MCRMST. RB(%): relative bias. SE: standard error. Coverage: coverage probability.

From Table 2, we can see that our MCRMST estimator is very close to true MCRMST, all the relative biases are smaller than 1%. The coverage probabilities are not far away from the nominal value. When u gets larger, the confidence intervals tend to have a larger coverage probability.

3.2. One arm with frailty

In this section, frailty is added to simulate the event correlation within subjects. The frailty is generated from Gamma distribution with variance 0.5. The shape parameters for the Weibull distributions are 15. Table 3 shows a similar conclusion with Table 2.

3.3. Two arms without frailty

In this subsection, we present the numerical results for two-arm case without frailty. Type-1 error and power are reported under null and alternative hypotheses, respectively. Under the alternative

Table 3. MCRMST: one arm with frailty.

t1	u	t2	nEvn	CR	DECR	Truth	Estimator	RB(%)	SE	Coverage
0	600	2930	245.84	0.58	0.95	2234.64	2233.93	−0.03	17.95	0.946
0	800	2930	342.71	0.58	0.95	2862.50	2861.64	−0.03	28.95	0.947
0	1000	2930	427.21	0.58	0.95	3428.07	3426.86	−0.04	41.60	0.948
0	1200	2930	494.94	0.58	0.95	3936.43	3934.61	−0.05	55.11	0.959
0	1400	2930	547.81	0.58	0.95	4394.09	4391.89	−0.05	69.41	0.966
0	1600	2930	588.72	0.58	0.95	4807.38	4804.77	−0.05	84.87	0.972
600	800	2930	97.36	0.59	0.95	3156.31	3156.51	0.01	5.12	0.958
600	1000	2930	180.14	0.59	0.95	3828.59	3829.39	0.02	14.42	0.948
600	1200	2930	248.88	0.59	0.95	4424.78	4426.35	0.04	26.46	0.946
600	1400	2930	302.51	0.59	0.95	4954.16	4955.99	0.04	40.18	0.944
600	1600	2930	343.10	0.59	0.95	5425.82	5427.95	0.04	55.30	0.949
600	1800	2930	373.32	0.59	0.95	5848.25	5850.43	0.04	71.88	0.959
1200	1400	2930	53.53	0.69	0.95	5544.44	5544.56	0.00	9.12	0.952
1200	1600	2930	93.95	0.69	0.95	6189.32	6189.50	0.00	24.79	0.968
1200	1800	2930	124.08	0.69	0.95	6751.02	6751.59	0.01	44.41	0.975
1200	2000	2930	146.12	0.69	0.95	7243.90	7244.01	0.00	67.55	0.985
1200	2200	2930	161.13	0.69	0.95	7679.77	7683.92	0.05	93.25	0.991

Notes. nEvn: number of events. CR: censor rate. DECR: death event censor rate. Truth: true MCRMST. RB(%): relative bias. SE: standard error. Coverage: coverage probability.

hypothesis, increase the scale parameters of the four Weibull distributions by 500 to construct the true mean survival time difference.

In Tables 4 and 5, 'nEvn1′ and 'nEvn2′ provide a number of events observed up to cut-off time 'u' for the two arms, respectively. Under the null hypothesis, 'Truth's are 0. Under the alternative hypothesis, 'Truth's are true MCRMST difference between the population of the two arms conditional on the interval of interest. 'DIFF' shows the difference of MCRMST estimator between the two arms $\hat{D}_{t_1,t_2}(u)$. 'SE' column shows the perturbation re-sampling standard error of $\hat{D}_{t_1,t_2}(u)$. 'Type-I' error and 'Power' are reported under null and alternative hypotheses, respectively. Besides, a time-to-first-event CRMST hypothesis test power is reported to compare with the power of the proposed test.

From Table 4, we can see that under the null hypothesis, our MCRMST difference estimators are quite close to 0. The standard error gets larger as u increases. The type-1 errors are close to 0.05. Under the alternative hypothesis, results in Table 5 show that 'Truth' is close to 'DIFF'. The power of our test is larger than those of the time-to-first-event CRMST hypothesis test.

3.4. Two arms with frailty

In this section, we present the numerical results for two-arm case with frailty, which is generated in the same way as stated in section 3.2. Type-1 error and power are reported under the null and alternative hypotheses, respectively. Under the alternative hypothesis, increase the scale parameters of the all four Weibull distributions by 300 to construct the true mean survival time difference. Tables 6 and 7 show similar conclusions as Tables 4 and 5.

3.5. Simulation summary

In this simulation, we studied the performance of MCRMST estimator in different scenarios. Across all scenarios, the relative biases are small and the coverage probabilities are close to the nominal level indicating that our MCRMST estimator has good finite sample properties, with or without frailty involved. Furthermore, in two-arm cases, the proposed hypothesis test has higher power and outperforms the time-to-first-event (single event) CRMST hypothesis test.

4. Real data

For illustration, we re-analyze the data from PEGASUS-TIMI 54 (Bonaca, 2015)(Bonaca et al. 2015) and focus only on the 60 mg treatment and placebo groups. The primary endpoint of interest is overall mean survival time with respect to CVD, MI, and STROKE. First, to make comparison with the traditional analyses, the study period of interest is set from randomization (Day 0, t_1) to Day 2000

Table 4. MCRMST: two arms without frailty under H_0.

t1	u	t2	nEvn1	nEvn2	CR	DECR	Truth	DIFF	SE	Type-I error
0	600	2930	456	456	0.54	0.80	0.00	0.02	26.74	0.063
0	800	2930	525	524	0.54	0.80	0.00	−0.01	39.52	0.052
0	1000	2930	576	576	0.54	0.80	0.00	0.01	53.55	0.037
0	1200	2930	616	616	0.54	0.80	0.00	0.05	68.92	0.029
0	1400	2930	647	647	0.54	0.80	0.00	−0.20	85.41	0.021
600	800	2930	68	69	0.67	0.80	0.00	−0.26	7.85	0.046
600	1000	2930	120	121	0.67	0.80	0.00	−0.71	21.67	0.039
600	1200	2930	161	161	0.67	0.80	0.00	−0.75	39.03	0.032
600	1400	2930	191	192	0.67	0.80	0.00	−0.11	59.27	0.023
600	1600	2930	215	216	0.67	0.80	0.00	0.54	82.50	0.017

Notes. nEvn1: number of events for arm1. nEvn2: number of events for arm2. CR: average censor rate of the two arms. DECR: average death event censor rate of the two arms. Truth: true MCRMST difference. DIFF: MCRMST estimator difference. SE: standard error of the difference estimator.

Table 5. MCRMST: two arms without frailty under H_1.

t1	u	t2	nEvn1	nEvn2	CR	DECR	Truth	DIFF	SE	Power	TTFE power
0	600	2930	455	379	0.57	0.81	81.62	81.19	25.74	0.867	0.778
0	800	2930	524	442	0.57	0.81	127.03	125.80	38.37	0.906	0.789
0	1000	2930	576	490	0.57	0.81	176.63	174.29	52.16	0.928	0.782
0	1200	2930	616	528	0.57	0.81	228.97	225.28	67.36	0.943	0.769
0	1400	2930	647	557	0.57	0.81	283.01	278.14	83.88	0.942	0.735
600	800	2930	69	62	0.68	0.81	8.62	8.63	7.24	0.215	0.088
600	1000	2930	121	110	0.68	0.81	30.99	30.98	20.06	0.330	0.113
600	1200	2930	162	148	0.68	0.81	63.69	63.66	36.42	0.413	0.131
600	1400	2930	193	178	0.68	0.81	103.98	103.95	55.57	0.458	0.133
600	1600	2930	217	202	0.68	0.81	149.69	149.41	77.49	0.486	0.109

Notes. nEvn1: number of events for arm1. nEvn2: number of events for arm2. CR: average censor rate of the two arms. DECR: average death event censor rate of the two arms. Truth: true MCRMST difference. DIFF: MCRMST estimator difference. SE: standard error of the difference estimator. TTFE Power: time-to-first-event test power.

Table 6. MCRMST: two arms with frailty under H_0.

t1	u	t2	nEvn1	nEvn2	CR	DECR	Truth	DIFF	SE	Type-I error
0	600	2930	247	248	0.55	0.95	0.00	−0.45	25.12	0.045
0	800	2930	348	349	0.55	0.95	0.00	−1.25	40.52	0.046
0	1000	2930	435	436	0.55	0.95	0.00	−2.19	57.75	0.042
0	1200	2930	506	507	0.55	0.95	0.00	−3.04	76.27	0.036
0	1400	2930	564	565	0.55	0.95	0.00	−3.89	95.88	0.029
600	800	2930	101	100	0.57	0.95	0.00	−0.02	7.16	0.037
600	1000	2930	187	186	0.57	0.95	0.00	0.44	20.07	0.048
600	1200	2930	258	257	0.57	0.95	0.00	0.72	36.38	0.052
600	1400	2930	315	314	0.57	0.95	0.00	1.22	54.90	0.057
600	1600	2930	359	358	0.57	0.95	0.00	1.72	75.42	0.054

Notes. nEvn1: number of events for arm1. nEvn2: number of events for arm2. CR: average censor rate of the two arms. DECR: average death event censor rate of the two arms. Truth: true MCRMST difference. DIFF: MCRMST estimator difference. SE: standard error of the difference estimator.

Table 7. MCRMST: two arms with frailty under H_1.

t1	u	t2	nEvn1	nEvn2	CR	DECR	Truth	DIFF	SE	Power	TTFE power
0	600	2930	248	183	0.58	0.95	47.22	47.22	23.61	0.505	0.354
0	800	2930	349	267	0.58	0.95	89.39	89.28	38.53	0.636	0.518
0	1000	2930	436	343	0.58	0.95	140.68	140.59	55.47	0.730	0.628
0	1200	2930	507	410	0.58	0.95	197.18	196.87	73.98	0.786	0.678
0	1400	2930	564	466	0.58	0.95	255.87	255.37	93.78	0.799	0.699
600	800	2930	101	84	0.59	0.95	11.28	11.41	6.58	0.402	0.256
600	1000	2930	188	159	0.59	0.95	41.23	41.55	18.67	0.605	0.396
600	1200	2930	260	226	0.59	0.95	83.97	84.55	34.20	0.690	0.469
600	1400	2930	317	281	0.59	0.95	134.63	135.74	52.15	0.732	0.515
600	1600	2930	361	327	0.59	0.95	189.17	190.98	72.10	0.747	0.519

Notes. nEvn1: number of events for arm1. nEvn2: number of events for arm2. CR: average censor rate of the two arms. DECR: average death event censor rate of the two arms. Truth: true MCRMST difference. DIFF: MCRMST estimator difference. SE: standard error of the difference estimator.

which ensures that patients at risk on Day 0 include all randomized patients. The target time points of interest, u, are from Day 300 to Day 1100 by an increase of 200 days. The Kaplan–Meier plot as in Figure 4 summarizes the data by event types. As shown in Table 8, by the landmark Day 1100, all-three (CVD, MI, and STROKE), first-of-three, CVD, and STROKE types of events reach the maximum

treatment benefit in RMST. For MI, the maximum treatment benefit in RMST is at Day 900. For CVD, before Day 900, the treatment benefit is noisy without clear direction of treatment benefit.

To reduce the impact of random noise on the analysis results and focus on patients who survive at least 800 days after randomization, we further analyze the data in a conditional setting of $t_1 = 800$ and $t_2 = 2,000$. The Kaplan–Meier plot as in Figure 5 summarizes the data by event types. The analysis results are listed in Table 9. For patients who survive longer than 800 days, the maximum treatment benefit in RMST for all-three event type is the total-event-free time of 65 days which is shorter than the 109 days for all patients at day 1100. The reason is that patients who survive longer than 800 days do not benefit from treatment in terms of MI-event-free and benefit less from treatment in terms of STROKE-event-free but benefit more from treatment with respect to CVD-event-free. Overall, all-three-event-free survival time assesses patients' disease burden more efficiently and yields more clinically meaningful results.

5. Discussion and conclusions

When using RMST, the target time point of interest, u, needs to be pre-specified. In addition, selection of time point of interest, u, has been described in publications. The RMST R package 'surRM2′ developed by Uno et al. (2017) (Uno et al. 2017) suggested that u needs to be smaller than the minimum of the largest observed event time in each of the two groups. Tian et al. (2018)(Tian et al. 2018) recommended specifying u as the minimum of the 95^{th} percentiles of observed minimum of event or censoring times across the treatment groups. The rationales for these recommendations are not clear. We use a simple example in the following (see Figure 6) to illustrate the reasons behind. Assume there are two survival profiles which are the same except that profile I has a censored event and profile II has a death event at the end of the follow-up, 10 months. Suppose u is specified as the last observed event time, 8 months, for Profile I, and 10 months for Profile II. Thus, for profile I, the area as highlighted with stars which is (10–8) times 0.2 (i.e., 0.4 months) in the left panel is not counted into the RMST. And the difference in RMST between the two profiles (II vs. I) is 0.4 months. That is, patients in profile II survive longer than patients in profile I on average. Actually, this is not true. Following Uno' method, when comparing the two profiles, u should be restricted to 8 months for both profiles and ignore information collected from 8 to 10 months. As such, there is no difference in RMST up to month 8. For our method, by assuming patients who are censored at 10 months have events, which can be considered as the 'worst-case' scenario, the area from 0 to 10 months and highlighted with stars is counted towards the RMST. Thus, there is no difference in RMST up to 10 months, but patients in profile I survive longer. As shown in the above example, our method by using horizontal calculation seems to have advantage over Uno' vertical method in the sense that our method does not need to add the restriction on specifying u.

In the previous section, the main objective of the primary analysis is to show the treatment benefit. The events included are relevant to treatment effectiveness in a concordant way. However, treatment effectiveness is only one of the key aspects of drug evaluation. Drug safety is another key aspect. Under the benefit-risk analysis framework, these two aspects are evaluated collectively either qualitatively or quantitatively. Our proposed MCRMST estimator can handle the efficacy and safety events jointly and be used for a quantitative benefit-risk analysis. Even though the perturbation re-sampling method is computation intensive in some sense, the running time per our experience is still tolerable. In addition, parallel computing techniques can further improve the computational efficiency.

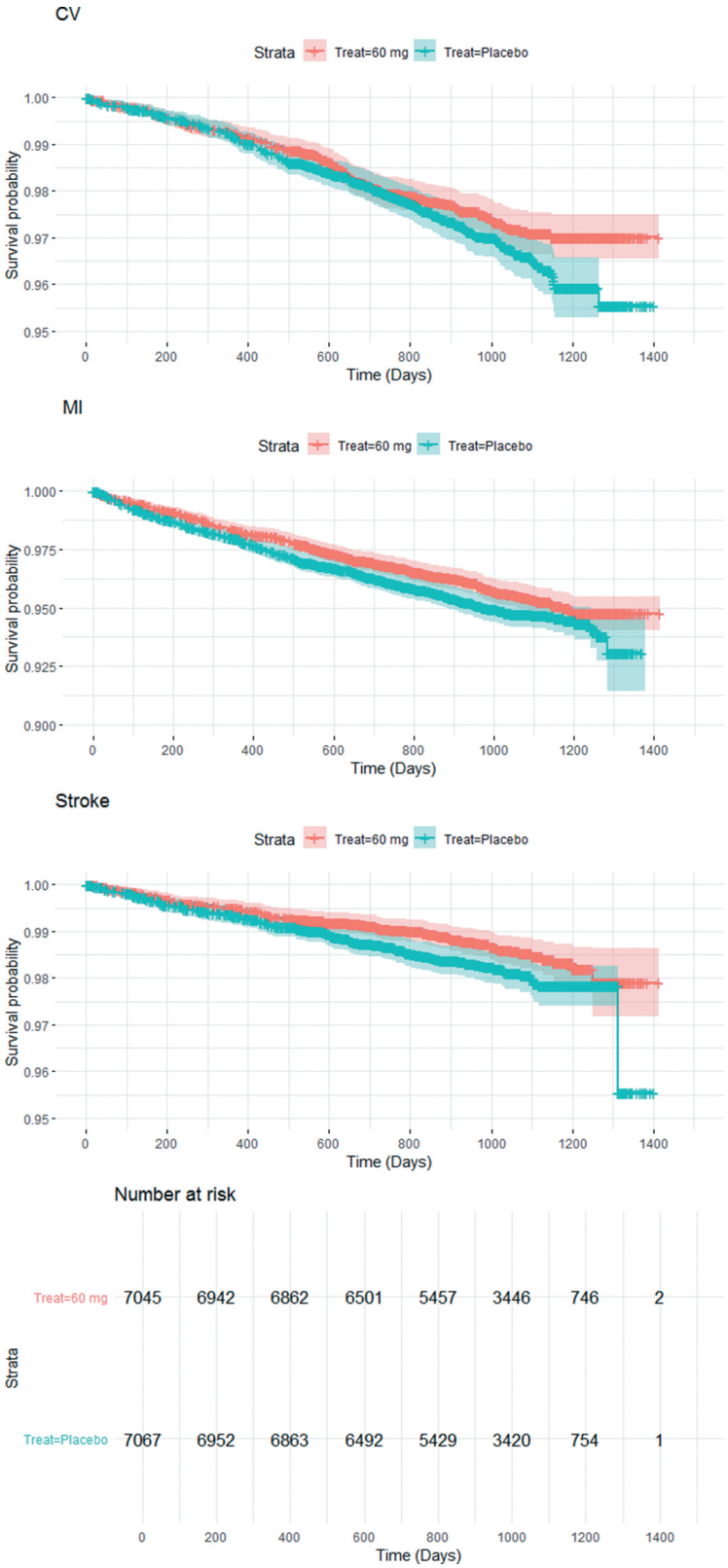

Figure 4. Kaplan–Meier Plot of PEGASUS Data by Clinical Events [0, 2000].

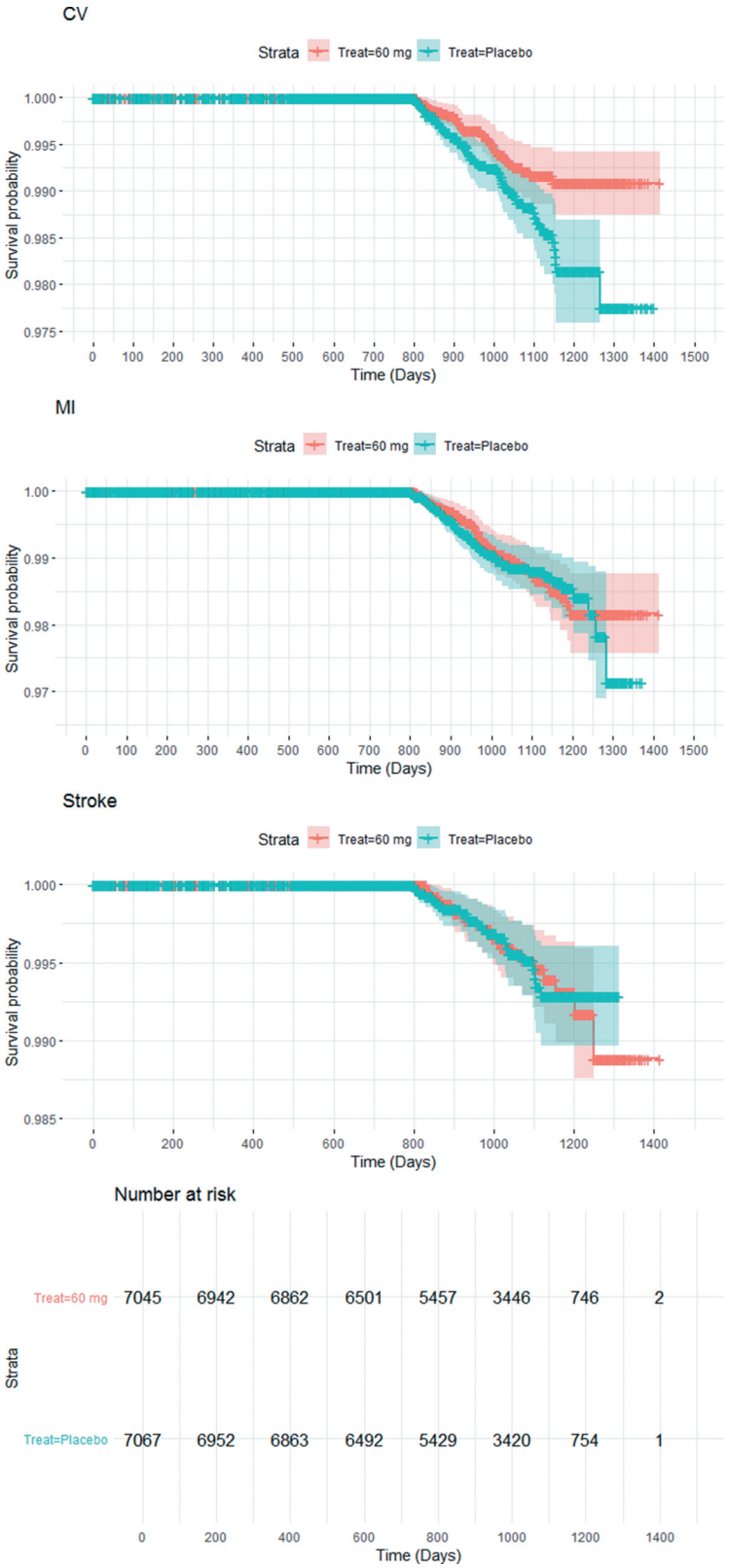

Figure 5. Kaplan–Meier Plot of PEGASUS Data by Clinical Events [800, 2000].

Table 8. MCRMST: PEGASUS [0, 2000].

Type	t1	u	t2	nE(60)	nE(P)	MCRMST (60 v.s. P)	SE	*p*-value
CVD + MI + STROKE	0	300	2000	176	213	3.49	0.48	< .0001
CVD + MI + STROKE	0	500	2000	281	359	9.44	1.03	< .0001
CVD + MI + STROKE	0	700	2000	399	473	14.32	1.70	< .0001
CVD + MI + STROKE	0	900	2000	474	582	77.50	2.53	< .0001
CVD + MI + STROKE	0	1100	2000	537	648	109.11	3.91	< .0001
CVD/MI/STROKE	0	300	2000	161	199	4.07	0.45	< .0001
CVD/MI/STROKE	0	500	2000	254	323	11.28	0.98	< .0001
CVD/MI/STROKE	0	700	2000	355	416	15.61	1.51	< .0001
CVD/MI/STROKE	0	900	2000	421	509	62.15	2.20	< .0001
CVD/MI/STROKE	0	1100	2000	476	563	62.38	3.11	< .0001
CVD	0	300	2000	46	45	0.31	0.24	0.1934
CVD	0	500	2000	78	97	1.34	0.59	0.0237
CVD	0	700	2000	129	130	−3.66	1.01	0.0005
CVD	0	900	2000	151	173	18.56	1.48	< .0001
CVD	0	1100	2000	173	200	42.14	2.35	< .0001
MI	0	300	2000	100	127	2.27	0.36	< .0001
MI	0	500	2000	153	199	6.04	0.73	< .0001
MI	0	700	2000	210	256	11.66	1.18	< .0001
MI	0	900	2000	248	303	44.38	1.69	< .0001
MI	0	1100	2000	278	330	27.98	2.52	< .0001
STROKE	0	300	2000	30	41	0.91	0.20	< .0001
STROKE	0	500	2000	50	63	2.06	0.42	< .0001
STROKE	0	700	2000	60	87	6.32	0.70	< .0001
STROKE	0	900	2000	75	106	14.56	1.08	< .0001
STROKE	0	1100	2000	86	118	38.99	1.79	< .0001

nE : number of events

Table 9. MCRMST: PEGASUS [800, 2000].

Type	t1	u	t2	nE(60)	nE(P)	MCRMST (60 v.s. P)	SE	*p*-value
CVD + MI + STROKE	800	900	2000	33	50	39.41	0.10	< .0001
CVD + MI + STROKE	800	1000	2000	73	90	16.80	0.35	< .0001
CVD + MI + STROKE	800	1100	2000	96	116	65.45	0.69	< .0001
CVD/MI/STROKE	800	900	2000	29	42	32.86	0.09	< .0001
CVD/MI/STROKE	800	1,000	2000	64	74	2.90	0.31	< .0001
CVD/MI/STROKE	800	1100	2000	84	96	33.77	0.58	< .0001
CVD	800	900	2000	10	21	19.26	0.06	< .0001
CVD	800	1000	2000	22	35	23.08	0.18	< .0001
CVD	800	1100	2000	32	48	44.33	0.43	< .0001
MI	800	900	2000	15	21	18.93	0.07	< .0001
MI	800	1000	2000	36	41	−0.70	0.25	0.0057
MI	800	1100	2000	45	48	−2.92	0.43	< .0001
STROKE	800	900	2000	8	8	1.22	0.05	< .0001
STROKE	800	1000	2000	15	14	−5.58	0.15	< .0001
STROKE	800	1100	2000	19	20	24.03	0.33	< .0001

nE : number of events

In conclusion, we propose an MCRMST estimator for evaluating time-to-event data with multiple types of clinical events, assess its variance using a perturbation re-sampling method, characterize its performance via comprehensive simulations. Overall speaking, the estimation procedure performs reasonably well.

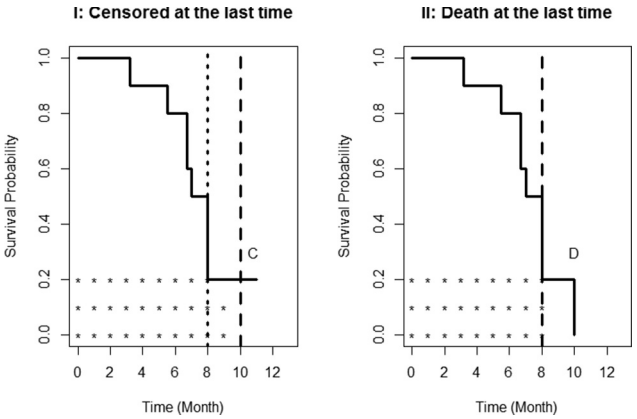

Figure 6. Two Survival Profiles.

Disclaimer

This paper reflects the views of the authors and should not be construed to represent FDA's views or policies.

References

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding. 2012. *Statistical models based on counting processes*. New York: Springer Science & Business Media.

Bonaca, M. P., D. L. Bhatt, M. Cohen, P. G. Steg, R. F. Storey, E. C. Jensen, G. Magnani, S. Bansilal, M. P. Fish, K. Im, et al. 2015. Long-term use of ticagrelor in patients with prior myocardial infarction. *New England Journal of Medicine*. 372(19):1791–1800. doi:10.1056/NEJMoa1500857.

Claggett, B., L. Tian, H. Fu, S. D. Solomon, and L.-J. Wei. 2018. Quantifying the totality of treatment effect with multiple event-time observations in the presence of a terminal event from a comparative clinical study. *Statistics in Medicine* 37 (25):3589–3598. doi:10.1002/sim.7907.

Gill, R. 1983. Large sample behaviour of the product-limit estimator on the whole line. *The Annals of Statistics* 11 (1):49–58. doi:10.1214/aos/1176346055.

Hougaard, P. 2019. *Draft qualification opinion of clinically interpretable treatment effect measures based on recurrent event endpoints that allow for efficient statistical analyses: Call for comments to ema*. European Medicine Agency, Netherlands.

Lin, D. Y., L.-J. Wei, I. Yang, and Z. Ying. 2000. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62 (4):711–730. doi:10.1111/1467-9868.00259.

Lin, D. Y., L.-J. Wei, and Z. Ying. 1993. Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika* 80 (3):557–572. doi:10.1093/biomet/80.3.557.

Michael Parzen, Z. Y., and L. Wei. 1997. Simultaneous confidence intervals for the difference of two survival functions. *Mathematics* 24 (3):309–314.

Qiu, J., E. Gu, D. Zhou, J. Lawrence, S. Bai, and H. M. J. Hung. 2019. Estimation on conditional restricted mean survival time with counting process. *Journal of Biopharmaceutical Statistics* 29 (5):800–809. doi:10.1080/10543406.2019.1657129.

Tian, L., H. Fu, S. J. Ruberg, H. Uno, and L.-J. Wei. 2018. Efficiency of two sample tests via the restricted mean survival time for analyzing event time observations. *Biometrics* 74 (2):694–702. doi:10.1111/biom.12770.

Uno, H., L. Tian, A. Cronin, C. Battioui, M. Horiguchi, and M. H. Uno. Package 'survrm2'. 2017.

Zhao, L., B. Claggett, L. Tian, H. Uno, M. A. Pfeffer, S. D. Solomon, L. Trippa, and L. Wei. 2016. On the restricted mean survival time curve in survival analysis. *Biometrics* 72 (1):215–221. doi:10.1111/biom.12384.

Zhao, L., L. Tian, H. Uno, S. D. Solomon, M. A. Pfeffer, J. S. Schindler, and L. J. Wei. 2012. Utilizing the integrated difference of two survival functions to quantify the treatment contrast for designing, monitoring, and analyzing a comparative clinical study. *Clinical Trials* 9 (5):570–577. doi:10.1177/1740774512455464.

Appendix.

Simplification in Calculation

The conditional estimator of RMST for a type k event by Equation (2) can be further simplified via assuming each subject, who is censored after the last observed event time and before t_2, which is usually the administrative censoring time (i.e., trial stopping time or end of study period), has an 'event' at $min(u, t_2)$ and contributes to the calculation of RMST by $min(u, t_2)$ from timescale. As such, $\hat{S}_k(t_2)$ is 0. The simplified equation is

$$\hat{m}^k_{t_1,t_2}(u) = \frac{1}{\hat{S}_k(t_1)} \sum_{i=1}^{n_k} \frac{\hat{S}_k(X_{ik-1})\Delta_{ik} I(t_1 \le X_{ik} < t_2)}{R(X_{ik-1})} \int_0^u I(X_{ik} \ge s) ds, \tag{4}$$

where $R(X_{ik-1}) = n_k * \bar{R}(X_{ik-1})$ is the number of subjects at risk at the time point X_{ik-1}. A product-limit estimate for type-k event is

$$\hat{S}_k(t_1) = \{1 - \frac{d_k(X_{1k})}{R_k(X_{1k})}\} \dots \{1 - \frac{d_k(X_{t_1k})}{R_k(X_{t_1k})}\}, \tag{5}$$

where $d_k(X_{t_nk})$ is the number of event k by time X_{t_nk}.

Similarly,

$$\hat{S}_k(X_{ik-1}) = \{1 - \frac{d_k(X_{1k})}{R_k(X_{1k})}\} \dots \{1 - \frac{d_k(X_{t_1k})}{R_k(X_{t_1k})}\} \dots \{1 - \frac{d_k(X_{t_1k+1})}{R_k(X_{t_1k+1})}\} \dots \{1 - \frac{d_k(X_{ik-1})}{R_k(X_{ik-1})}\}. \tag{6}$$

As such,

$$\frac{\hat{S}_k(X_{ik-1})}{\hat{S}_k(t_1)} = \{1 - \frac{d_k(X_{t_1k+1})}{R_k(X_{t_1k+1})}\} \dots \{1 - \frac{d_k(X_{ik-1})}{R_k(X_{ik-1})}\} = \hat{S}^{t1}_k(X_{ik-1}), \tag{7}$$

which is the survival function estimated based on the observations after t_1.

Therefore, Equation (6) can be further simplified as

$$\hat{m}^k_{t_1,t_2}(u) = \sum_{i=1}^{n_k} \frac{\hat{S}^{t_1}_k(X_{ik-1})\Delta_{ik} I(t_1 \le X_{ik} < t_2)}{R(X_{ik-1})} \int_0^u I(X_{ik} \ge s) ds \tag{8}$$

As demonstrated earlier, by Uno' method, each subject who is censored after the last observed event time (t_{last}) can only contribute by a maximum of t_{last} given $u \le t_{last}$ from the timescale. For our method, those subjects can contribute to RMST calculation beyond t_{last} through the 'worst-case' assumption.